PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2001-242890
(43)Date of publication of application : 07.09.2001

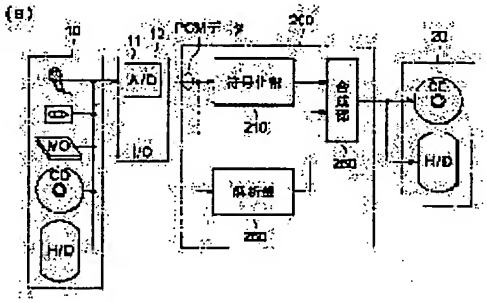
(51)Int.Cl. G10L 19/00
G09B 5/04
G11B 20/02
G11B 20/10
H04N 5/92
// H04N 7/173

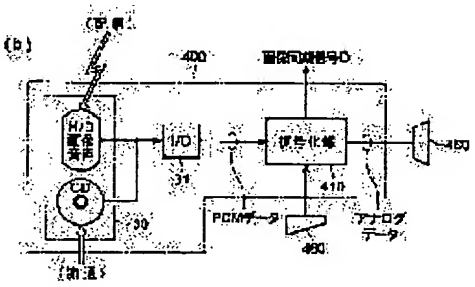
(21)Application number : 2000-051801
(22)Date of filing : 28.02.2000

(71)Applicant : KANAASU DATA KK
(72)Inventor : SEKIGUCHI HIROSHI

(54) DATA STRUCTURE OF VOICE DATA, GENERATING METHOD, REPRODUCING METHOD, RECORDING METHOD, RECORDING MEDIUM, DISTRIBUTION METHOD AND REPRODUCING METHOD OF MULTIMEDIA

(57)Abstract:
PROBLEM TO BE SOLVED: To provide a data structure of voice data in which voice data having a changed reproducing speed are decoded at a user's side without degrading listening quality.
SOLUTION: A synthesis section 260 adds decoding auxiliary information, which includes kinds of sound of various sections constituting generated sound to be referred while decoding coded voice data specified by an analysis section 250, to the coded voice data from voice signals coded by a coding section 210 in accordance with a prescribed rule. Thus, decoding of voice data having an arbitrary reproducing speed is accomplished at the user's side and the voice data are made suitable for the contents of data distribution services using information communication technology.





LEGAL STATUS

[Date of request for examination] 07.03.2002

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision

of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

CLAIMS

[Claim(s)]
[Claim 1] The data structure of the voice data equipped with the decryption auxiliary information which is information referred to in the case of a decryption of the coding voice data encoded according to the predetermined rule from the sound signal, and the aforementioned coding voice data, and includes the information about the kind of sound of each part which constitutes generating sound at least specified from the physical quantity about the wave motion of the aforementioned sound signal.
[Claim 2] The data structure of the voice data according to claim 1 characterized by including the frequency spectrum information on the aforementioned sound signal in the physical quantity about the wave motion of the aforementioned sound signal.
[Claim 3] The data structure of the voice data according to claim 1 or 2 characterized by containing in the aforementioned decryption auxiliary information the emphasis position identification information for directing the position of the direction of a time-axis which should be emphasized in the amplitude direction.
[Claim 4] The data structure of the voice data of the claim 1-3 characterized by including the information which directs individually the frequency component which should be emphasized among the frequency components of the voice data by which the decryption was carried out [aforementioned] in the aforementioned decryption auxiliary information given in any 1 term.
[Claim 5] The data structure of the voice data of the claim 1-4 characterized by including the information which directs the display timing of the image data which should be displayed on a predetermined display means given in any 1 term.
[Claim 6] The generation method of voice data characterized by providing the following About the 1st line that generates the coding voice data encoded according to the predetermined rule from the sound signal About the 2nd line that specifies the information about the kind of sound of each part which constitutes generating sound from physical quantity about the wave motion of the aforementioned sound signal at least as decryption auxiliary information referred to in the case of a decryption of the aforementioned coding voice data The 3rd process which generates synthetic data new as the aforementioned voice data by adding the aforementioned decryption auxiliary information specified as the aforementioned coding voice data generated about in the 1st aforementioned line about in the 2nd aforementioned line
[Claim 7] The reproduction method of the voice data which reproduces voice at the speed adjusted for every sound of each part which constitutes generating sound based on the reproduction-speed information which is characterized by providing the following, and which was beforehand specified in the voice data which has a data structure according to claim 1 About the 1st line that extracts the decryption auxiliary information referred to in the case of a decryption of the aforementioned voice data to the aforementioned coding voice data the reproduction speed which is suitable for voice reproduction for every sound of each part which constitutes this generating sound contained in the aforementioned coding voice data while referring to the information about the sound of each part which is contained in the decryption auxiliary information extracted about in the 1st aforementioned line, and which constitutes generating sound at least -- the account of before -- about the 2nd line that determines on the basis of

the reproduction-speed information specified beforehand About the 3rd line that decrypts this coding voice data while performing extension processing or shortening processing to the applicable portion of this coding voice data so that it may be equivalent to the reproduction speed by which a decision was made [aforementioned] for every sound of each part which constitutes the generating sound contained in the aforementioned coding voice data

[Claim 8] The record method of the voice data which records the decryption auxiliary information which is information referred to in the case of a decryption of the coding voice data encoded according to the predetermined rule from the sound signal, and the aforementioned coding voice data, and includes the information about the kind of sound of each part which constitutes generating sound at least specified from the physical quantity about the wave motion of the aforementioned sound signal on a predetermined record medium.

[Claim 9] The record medium of voice data with which the decryption auxiliary information which is information referred to in the case of a decryption of the coding voice data encoded according to the predetermined rule from the sound signal and the aforementioned coding voice data, and includes the information about the kind of sound of each part which constitutes generating sound at least specified from the physical quantity about the wave motion of the aforementioned sound signal was recorded.

[Claim 10] The distribution method of the voice data which transmits the decryption auxiliary information which is information referred to in the case of a decryption of the coding voice data encoded according to the predetermined rule from the sound signal, and the aforementioned coding voice data, and includes the information about the kind of sound of each part which constitutes generating sound at least specified from the physical quantity about the wave motion of the aforementioned sound signal to a communications partner through the means of communication of information of a cable or radio.

[Claim 11] Once develop 1 or the image data beyond it on memory, and the image data for one frame is synchronized with reproduction operation of the voice data which has a data structure according to claim 1 among the image data stored on this memory. It is the reproduction method of the multimedia displayed on a predetermined display means one by one. One frame or the criteria rewriting period of the N times (positive rational number) as many image data as this among two or more aforementioned image data stored in the aforementioned memory Tv, When making the reproduction time period determined for every sound of Ta and each part which constitutes generating sound based on the directed reproduction-speed information in the criteria reproduction time period determined based on the audio digitization sampling period into Ta' (>Ta), The reproduction method of the multimedia which stops next image data rewriting operation from the end point in time of image data rewriting operation to predetermined timing to Tvx (Ta'/Ta) (-1).

[Claim 12] Once develop 1 or the image data beyond it on memory, and the image data for one frame is synchronized with reproduction operation of the voice data currently recorded on the record medium of voice data according to claim 1 among the image data stored on this memory. the criteria reproduction time period which is the reproduction method of the multimedia displayed on a display means one by one, and was determined based on the audio digitization sampling period -- Ta -- and When making the reproduction time period determined for every sound of each part which constitutes generating sound based on the directed reproduction-speed information into Ta' (>Ta), The reproduction method of the multimedia which was beforehand specified in the average rewriting frequency of the image data stored in the aforementioned memory and which is rewritten and is set up the twice (Ta'/Ta) of frequency.

[Translation done.]

DETAILED DESCRIPTION

[Detailed Description of the Invention]
[0001]
[The technical field to which invention belongs] This invention relates to the data structure of the voice data which enables a decryption of the voice data for reproduction by which the reproduction speed was arbitrarily changed by the user side, a generation method, the reproduction method, the record method, a record medium, the distribution method, and the reproduction method of the multimedia which synchronized with the reproduction speed of this voice data, without spoiling the ease of catching.
[0002]
[Description of the Prior Art] From the former, the teaching materials with which speech information was recorded on record media, such as a cassette tape, with music for the object for self-study of language studies, such as English conversation, the object for practice of ****, the object for legal self-study, practice of a song, and the other purposes are offered variously. Here, when the teaching materials for self-study of English conversation were explained to the example, the conventional main record media are the cassette tapes (or CD) with which phonation (speech information) of a series of English was recorded, and the student was using it combining this tape teaching materials and text. In addition, various level is prepared for such teaching materials from the object for the beginners' classes to the object for upper classes.
[0003] Moreover, the 1st field where the speech information train (sound of each part which constitutes the generating sound of natural speed) suitable for the upper person study classified into two or more partitions was recorded on the Japan patent No. 2581700, The 2nd field where the speech information train (voice from which it is the sound of each part which constitutes the generating sound carried out clearly, and a descendant differs in a meaning with the same linguistics top) suitable for the beginners' class person study which consists of an equivalent partition corresponding to each [these] partition was recorded, The relation of each partition to which this object for upper person study and each speech information train for beginners' class person study correspond The reproduction method including the change reproduction during the partition to which information record media, such as CD-ROM equipped with the 3rd field where the information shown in the record position in the record medium of each partition of these speech information train was recorded at least, and the information record medium equipped with such structure correspond etc. is proposed.
[0004]
[Problem(s) to be Solved by the Invention] As mentioned above, the sound of each part which constitutes a native speaker's generating sound is recorded on the 1st field on this medium by the information record medium of the Japan patent No. 2581700, and, in the 2nd field, the speech information train which consisted of pronunciation which ****(ed) in the same meaning is recorded on it by the language top. Therefore, when reproduction sound is not able to be caught to the midst by which the speech information train recorded on the 1st field is reproduced, By changing the speech information train (correspondence with the partition under reproduction of the 1st speech information train and the partition which should reproduce the 2nd speech information train being recorded on the 3rd field) of the same content recorded on the 2nd field, and reproducing, a student can recognize the voice which was not able to be caught. Moreover, if spread and highly efficient-

ization of information management systems, such as a personal computer in recent years, are taken into consideration, it is not impossible to generate the speech information train recorded on the 2nd field which ****(ed) from the speech information train recorded on the 1st field of the above because of shortening of work time, or curtailment of work cost, either.

[0005] However, a user's ease of catching will be spoiled in having only expanded a native speaker's voice uniformly along with the time-axis. That is, even if it is the case where only reincarnate Japanese people slowly and they enable it to mainly hear English of natural speed, it is because voice reproduction time may be lengthened simply and uniformly about each frequency component or time change of the sound of each part which having been shortened is inadequate and constitutes generating sound, for example, child Otobe's spectrum, may mean another sound as sound on language. for example, the consonant of pronunciation which the spectrum itself has the almost same type only by the pronunciation of bus-available (**) and PA (**) having the former quick spectrum change, and the latter being late, and is called bus-available (**) -- if time also including the section is elongated -- PA (**) and ******** -- it becomes things

[0006] On the other hand, those who cannot fully catch the voice reproduction speed recorded on the 2nd field of the above, either, those who cannot be satisfied at the offered natural speed are various, and a student's hearing level must also prepare beforehand two or more kinds of speech information according to each student's hearing level, if it is going to satisfy the student of different hearing level in this way separately. However, since a limitation is in the storage capacity of record media, such as CD, it is not realistic to prepare two or more kinds of speech information which could not choose the speech information which suited its hearing level by the student side in the present condition, and suited each student's hearing level.

[0007] Furthermore, the data distribution using computer networks, such as the Internet, also attracts attention by development of information communication technology in recent years. When considering offer of the speech information using such data distribution, it cannot be said that practical use level is reached in respect of communication time or communication cost still more for transmitting a lot of data.

[0008] Without having been made in order that this invention might solve the above technical problems, and spoiling the ease of catching The data structure of the voice data which enables a decryption of the voice data for reproduction of the reproduction speed which he wishes by the user side, The reproduction method of the voice data for decrypting the new voice data for reproduction of reproduction speed which he wishes by the user side from a generation method and this voice data, The record method of the voice data for recording this voice data on a predetermined record medium, The distribution method of the voice data for providing a user with the record medium and this voice data on which this voice data was recorded using a computer network or sanitation communication system, And it aims at offering the reproduction method of the multimedia for making possible image display which synchronized with reproduction operation of this voice data.

[0009]

[Means for Solving the Problem] In order to solve an above-mentioned technical problem, the data structure of the voice data concerning this invention is equipped with the coding voice data encoded according to the predetermined rule from the sound signal, and the decryption auxiliary information referred to in the case of a decryption of this coding voice data.

[0010] Especially the above-mentioned decryption auxiliary information includes the information about the kind of sound of each part which constitutes generating sound at least specified, the physical quantity about the wave motion, for example, the frequency spectrum information etc., on a sound signal etc. This is for canceling the fault which the sound which changes with above child Otobe's change hears. a consonant -- if the elongatedness of the section is stopped to bus-available (**) and a ******** limitation and only the vowel section is elongated or shortened at the voice reproduction time of a wish -- with bus-available (**) -- ******** -- it becomes things However it may elongate or shorten the vowel section, it can be set as the length (reproduction time of a wish) of a wish from ******** with the vowel.

[0011] It is required for linguistic study to, also make the people in the non-English area emphasize and hear alternatively only the voice and the specific frequency component over which it passes and which are hard to catch double precision and 3 times on the other hand. [being weak] In having

emphasized also including the vowel section, the whole becomes large too much and it is ineffective. You surely have to emphasize alternatively. Then, as for the above-mentioned decryption auxiliary information, it is desirable that the emphasis position identification information for directing the position which should be emphasized is included. Moreover, you may make it this decryption auxiliary information include the information which directs individually the frequency component which should be emphasized among the frequency components of the decrypted voice data.

[0012] In addition, in order to also make a different reproduction speed easy to catch as coding for obtaining the above-mentioned coding voice data, the sound signal for coding is beforehand decomposed into a frequency component, and this divided coding given in Japanese Patent Application No. No. 249672 [ten to] which data-izes the amplitude information etc. for every frequency component is suitable. Moreover, the sound signal which is a candidate for coding may be a digitized electrical signal, and may be any of the digital speech information recorded on the analog speech information read from the analog speech information incorporated through the microphone as the information source, the magnetic tape, etc., MO and CD, the hard disk, etc. However, A/D conversion of the case of analog speech information once needs to be carried out. Moreover, when compression coding of the data recorded on CD etc. is carried out, it is necessary to elongate this compressed data (defrosting).

[0013] as one of the information offer services for which information offer service according paying one's attention to fields, such as computer networks, such as the Internet which began to spread, a cable TV network, and health communication, in recent years to multimedia, such as an alphabetic data, voice data, still picture data, and a video data, was also widely offered, and used such information communication technology, in order to make this invention apply, regulation of the display timing of image data becomes indispensable Then, the image display (the animation is displayed especially) synchronized with reproduction operation of voice data becomes possible by including the information which directs the display timing of the image data which should be displayed on the voice data equipped with the above data structures by the predetermined display means.

[0014] Moreover, the generation method of the voice data concerning this invention is equipped with about the 1st line that generates the above-mentioned coding voice data, about the 2nd line that specifies the above-mentioned decryption auxiliary information, and about the 3rd line that adds decryption auxiliary information to this coding voice data. About by the 1st above-mentioned line, coding of a sound signal is performed according to a predetermined rule like the coding technology indicated by Japanese Patent Application No. No. 249672 [ten to], for example. About by the 2nd above-mentioned line, the information about the kind of sound of each part which constitutes generating sound from physical quantity (for example, frequency spectrum information) about the wave motion of a sound signal at least as decryption auxiliary information referred to in the case of a decryption of coding voice data is specified. In addition, it is also as possible as the above 1st and the 2nd line to carry out in parallel.

[0015] By using the voice data (coding voice data being included) generated as mentioned above, a decryption of the voice data for reproduction suitable for each student's hearing level adjusted for every sound of each part which constitutes the generating sound based on the specified reproduction-speed information is attained in a decryption of this voice data. Thus, by reproducing the decrypted voice data for reproduction, a student can hear the reproduction voice adjusted by the speed which self specified. Namely, the reproduction method of the voice data concerning this invention About the 1st line that extracts decryption auxiliary information from the voice data generated as mentioned above, Referring to the information about the sound of each part which is contained in the extracted decryption auxiliary information which was extracted and which constitutes generating sound at least Performing extension processing (reproduction of loose voice sake) shortening processing (reproduction of early voice sake) to the applicable portion of coding voice data so that it may be equivalent to the reproduction speed determined about as the 2nd line which determines a reproduction speed for every sound of each part which constitutes generating sound It has about the 3rd line that decrypts this coding voice data.

[0016] In addition, since the reproduction method concerned can prepare two or more kinds of voice data for reproduction from which a reproduction speed differs, without spoiling the ease of catching,

study while performing change reproduction between the voice data for reproduction from which the generated reproduction speed differs is also attained. Moreover, the break (pause) of the shift section which appears among child Otobe, and these vowel sections and child Otobe who appears before and after the vowel section in a voice spectrum and this vowel section, and voice etc. is contained in the sound of each part which constitutes the above-mentioned generating sound.

[0017] In addition, offer of the voice data equipped with the above data structures can consider the case where it is provided for a user with the gestalt once recorded on record media, such as CD, and the case where it is provided for a user through information means of communications. Even when using information communication technology, the momentary record to a hard disk etc. is indispensable, and as for the handling of voice data, decryption auxiliary information is recorded on a predetermined record medium with coding voice data, without spoiling the ease of catching by the record method of the voice data concerning this invention so that a decryption of the voice data for reproduction by which the reproduction speed was changed by the user side may be attained. In addition, in the record medium of the voice data obtained by the above record method, the field where coding voice data is recorded may differ from the field where decryption auxiliary information is recorded.

[0018] As the data distribution method transmitted to a communications partner through the means of communication of information of a cable or radio, the voice data equipped with the above data structures by the distribution method of the voice data concerning this invention The coding voice data encoded according to the predetermined rule from the sound signal, It is the information referred to in the case of a decryption of this coding voice data, and decryption auxiliary information including the information about the kind of sound of each part which constitutes generating sound from physical quantity about the wave motion of a sound signal at least is transmitted to a communications partner. In addition, you may be the composition of transmitting separately the coding voice data transmitted and decryption auxiliary information to a communications partner by the distribution method of the voice data concerning this invention.

[0019] In addition, when making reproduction operation of the voice data equipped with the above data structures correspond to multimedia reproduction of an alphabetic data, image data (a still picture and animation), etc., it is important to take the synchronization with animation reproduction and reproduction of the voice by which especially the reproduction speed was changed freely. That is, although the animation displays the image data of about 20 frames on the display one by one in 1 second, if the synchronization with display timing with audio reproduction operation cannot be taken, it will become an unnatural display action.

[0020] Then, the reproduction method of the multimedia concerning this invention once develops 1 or the image data beyond it on memory, and is characterized by displaying the image data for one frame on a predetermined display means one by one among the image data stored on this memory synchronizing with reproduction operation of the voice data equipped with the above data structures.

[0021] Concretely the reproduction method of the multimedia concerned the criteria reproduction time period determined based on Tv and the audio digitization sampling period in one frame or the criteria rewriting period of the N times (positive rational number) as many image data as this among two or more image data stored in memory -- Ta -- and When making the reproduction time period determined for every sound of each part which constitutes generating sound based on the directed reproduction-speed information into Ta' (>Ta), Next image data rewriting operation is stopped from the end point in time of image data rewriting operation to predetermined timing to Tvx (Ta'/Ta) (-1).

[0022] In addition, regulation of the display timing of image data When making the reproduction time period determined for every sound of Ta and each part which constitutes generating sound based on the directed reproduction-speed information in the criteria reproduction time period determined based on the audio digitization sampling period into Ta' (>Ta), About the average rewriting frequency of the image data stored in memory, even if it rewrites and makes it set up the twice (Ta'/Ta) of frequency, the image display which was specified beforehand and which synchronized with reproduction operation of voice data becomes possible.

[0023]

[Embodiments of the Invention] Hereafter, each operation gestalt, such as a data structure of the

voice data concerning this invention, is explained using drawing 1 - drawing 13 . In addition, the explanation which gives the same sign to the same portion and overlaps in explanation of a drawing is omitted.

[0024] The voice data equipped with the data structure concerning this invention makes it possible to decrypt new voice data for reproduction of the reproduction speed which the user set up freely by this user side, without spoiling the ease of catching at the time of reproduction. Such a use gestalt of voice data can consider various modes by development of digital technology in recent years, or maintenance of data communication environment. Drawing 1 is a conceptual diagram for explaining what industrial up use of the voice data equipped with the data structure concerning this invention is carried out.

[0025] As shown in drawing 1 (a), as the information source 10 used for generation of the voice data equipped with the data structure concerning this invention For example, the analog speech information which was incorporated directly or was already recorded on the magnetic tape etc. through the microphone, The digital speech information currently furthermore recorded on MO, CD (DVD is included), H/D (hard disk), etc. can be used, and the speech information specifically offered by the teaching materials marketed, a television station, a radio station, etc. can also be used. An editor 100 generates the voice data equipped with the data structure concerning this invention with voice data generation equipment 200 using such the information source 10. In addition, if the present data offer method is considered in this case, a user will be provided with the generated voice data in many cases in the state where it was once recorded on the record media 20, such as CD (DVD is included) and H/D. Moreover, when the image data related with the voice data concerned is recorded on these CDs or H/D, it fully thinks.

[0026] The voice data generated by the above-mentioned voice-data generation equipment 200 is the information referred to in the case of a decryption of the coding voice data encoded according to the predetermined rule from the digital sound signal taken out from the above-mentioned information source 10, and this coding voice data, and is the new voice data equipped with decryption auxiliary information including the information about the kind of sound of each part which constitutes generating sound at least specified from the physical quantity about the wave motion of this sound signal. In addition, since a different reproduction speed is also made easy to catch as coding for obtaining coding voice data, for example, the sound signal for coding can be beforehand decomposed into a frequency component, and this divided coding given in Japanese Patent Application No. No. 249672 [ten to] which data-izes the amplitude information etc. for every frequency component can be used. The coding voice data and the decryption auxiliary information which were generated are stored in a record medium 20 by voice data generation equipment 200. Thereby, multimedia, such as image data and an alphabetic data, is recorded on record media, such as CD, DVD, and H/D, with above-mentioned coding voice data and decryption auxiliary information.

[0027] As for especially CD and DVD as a record medium 20, it is common to be sold at a store like computer software, Music CD, etc. in being provided for a user as an appendix of a magazine (circulation in a commercial scene). Moreover, the generated voice data does not ask a cable and radio from a server 300, but when distributing to a user through information means of communications, such as the Internet and sanitation communication, it is fully considered.

[0028] In data distribution, the voice data generated by the above-mentioned voice data generation equipment 200 is once accumulated with image data etc. at the storage 310 (for example, H/D) of a server 300. And the once accumulated voice data is transmitted to H/D310 through a transmitter-receiver 320 (I/O in drawing) at a user terminal 400. In a user-terminal 400 side, the voice data received through the transmitter-receiver 450 is once stored in H/D (contained in external storage 30). On the other hand, in data offer using CD, DVD, etc., it is used as an external recording device 30 of this terminal unit by equipping CD drive and a DVD drive of a terminal unit 400 with CD which the user purchased.

[0029] Usually, the terminal unit 400 by the side of a user is equipped with the display 470 of an input unit 460, CRT, liquid crystal, etc., and the loudspeaker 480, and a loudspeaker 480 is ****** (ed) once the voice data currently recorded on external storage 300 with image data etc. is decrypted by the voice data of the reproduction speed which the user itself directed by the decryption section

410 (realization also by software is possible) of the terminal unit 400 concerned. On the other hand, the image data stored in external storage 300 is displayed on a display 470 the whole frame, once it is developed by VRAM432 (bit mapped display). In addition, if two or more kinds of voice data for reproduction from which a reproduction speed differs is prepared in this external storage 30 by accumulating the voice data for reproduction decrypted by the decryption section 410 one by one in the above-mentioned external storage 30, the change reproduction between two or more kinds of voice data from which a reproduction speed differs using the technology indicated by the Japan patent No. 2581700 will be attained by the user side.

[0030] As shown in drawing 1 (b), while a user displays the picture 471 related on a display 470, the voice outputted from a loudspeaker 480 will be heard. Under the present circumstances, the display timing of a picture may shift in the reproduction speed having been changed only for voice. Then, it is desirable to add beforehand the information which directs image display timing to the coding voice data generated in the above-mentioned voice data generation equipment 200 so that the decryption section 410 can control the display timing of image data.

[0031] Next, the detailed structure of the voice data generation equipment 200 shown in drawing 1 (a) and a voice data regenerative apparatus (terminal unit 400) is explained using drawing 2 . In addition, drawing 2 (a) is drawing showing the composition of voice data generation equipment 200, and drawing 2 (b) is drawing showing the composition of the terminal unit 400 as a voice data regenerative apparatus.

[0032] As shown in drawing 2 (a), the sound signal incorporated by voice data generation equipment 200 is offered from the information source 10. In addition, since both the speech information incorporated from a microphone among the speech information offered from this information source 10 and the speech information from a magnetic tape are analog voice data, before being inputted into the voice data generation equipment 200 concerned, they are changed into PCM data by A/D converter 11 (contained in I/O12). Moreover, the speech information already stored in MO, CD (DVD ****), and H/D is incorporated by the voice data generation equipment 200 concerned through I/O12 as PCM data. When the incorporated voice data is compressed, it is necessary to once thaw software etc.

[0033] The coding section 210 which generates the coding voice data encoded according to the predetermined rule from the sound signal (electrical signal) from the information source 10 by which voice data generation equipment 200 was pretreated as mentioned above, As decryption auxiliary information referred to in the case of a decryption of this coding voice data The analysis section 250 which specifies the information about the kind of sound of each part which constitutes generating sound from physical quantity (for example, frequency spectrum information) about the wave motion of a sound signal at least, It has the synthetic section 260 which adds the decryption auxiliary information specified as the coding voice data encoded by the coding section 210 by the analysis section 250. The coding voice data and the decryption auxiliary information which were outputted from this synthetic section 260 are recorded on the record media 20, such as CD, DVD, and H/D. In addition, the above-mentioned coding voice data and decryption auxiliary information may be recorded on the field to which it differs in a record medium 20, respectively.

[0034] On the other hand, in a user side, as shown in drawing 2 (b), the voice data offered with the gestalt of data distribution, CD, etc. is stored in the external storage 30 of a terminal unit 400. The decryption section 410 outputs the picture synchronizing signal D while decrypting the digital data read from external storage 30 through I/O31 according to a user's content of directions inputted through the input meanses 460, such as a keyboard and pointing devices, such as a mouse, as voice data for reproduction reproducible at the rate of predetermined. After the decrypted voice data for reproduction is changed into analog data, it is outputted as voice from a loudspeaker 480.

[0035] In addition, the above-mentioned decryption section 410 reads the voice data read from external storage 30 through I/O31. While extracting the decryption auxiliary information referred to in the case of a decryption of this voice data to the read coding voice data The reproduction speed which was suitable for voice reproduction for every sound of each part which constitutes this generating sound contained in the above-mentioned coding voice data while referring to the information about the sound of each part which is contained in the extracted decryption auxiliary information, and which constitutes generating sound at least is determined on the basis of the

reproduction-speed information specified by the user. A decryption of the coding voice data in this -- decryption section 410 is performed, performing extension processing or shortening processing to the applicable portion of this coding voice data so that it may be equivalent to the reproduction speed determined as mentioned above for every sound of each part which constitutes the generating sound contained in this coding voice data.

[0036] Drawing 3 is drawing showing the structure of the coding section 210 in above-mentioned voice data generation equipment 200. The coding section 210 incorporates the sound signal which is equivalent to the voice of a native speaker's natural speed sampled by sound clock 44.1kHz of for example, the music CD with a microphone etc. first. this incorporated sound signal -- once -- each -- it is filtered in order to divide into channel CH#1-CH#85 (frequency component) In addition, the frequency range of the incorporated sound signal is 75Hz - 10,000Hz, and a sampling frequency is 44.1kHz (22.68 microseconds) in all at the sound clock of Music CD. the number of channels to divide -- 85 (7 octave +1 sound) -- carrying out -- each -- the center frequency (center f) of channel #1-#85 -- the semitone of temperament (it considers as an equal temperament of 12 degrees per octave) -- it is set up so that it may become a train (77.78Hz (D#) - 9,960Hz (D#))

[0037] above -- each -- as for the data divided into channel #1-#85, respectively, the amplitude information is extracted every (when one wave cannot be formed in 100 data of a 44.1kHz sampling by considerable, however 100 data, the number of data is increased) 2.268ms therefore -- this operation gestalt -- each -- the sampling rates (the 2nd period) of the amplitude information in channel #1-#85 are 441 samples /s (2.268ms) in addition, incorporating a sampling rate by 120 data to the degree incorporated by 100 data, and processing it to it that what is necessary is just a regular period, etc. -- ******** -- you may be the operation gestalt which repeats processing by turns at a rate

[0038] the coding section 210 was sampled every 2.268ms -- each -- the amplitude information on channel #1-#85 is expressed by 1 byte (8 bits), respectively, and 85 bytes (85 channels x 1 byte) of coding voice data a1, a2, a3, --, an is generated In addition, in order to control display timing with the dynamic image displayed at the time of reproduction of the voice equivalent to this coding voice data, the picture synchronizing signal D (1 byte) is added to the coding voice data a1, a2, a3, --, an.

[0039] On the other hand, the analysis section 250 of the voice data generation equipment 200 shown in drawing 3 specifies the decryption auxiliary information referred to in the case of a decryption of the coding voice data generated in the above-mentioned coding section 210.

[0040] The sound of each part which constitutes generating sound from spectrum information on the incorporated sound signal, the emphasis position identification information which shows the part which should be emphasized, the frequency component which should be emphasized, etc. are contained in decryption auxiliary information.

[0041] In this operation gestalt, for example, the sound of each part which constitutes generating sound With child Otobe who appears forward and backward on both sides of [as shown in drawing 4] the vowel section (V) and this vowel section (V) (CF and back child Otobe are shown for former child Otobe by CR among drawing) It is classified into the pause (P) which indicates the silent period which appears between each voice to be the shift section (for TF and the post-shift section to be shown for the pre-shift section by TR among drawing) which appears between the vowel section (V) and each former child Otobe CF and CR. In addition, a pause (P) may spoil a user's ease of catching in having been extended like the sound of each part which constitutes other generating sound in the case of delay of reproduction voice. Then, when generating in this operation gestalt between the case (P1) where generating a pause (P) between syllable is shown, the case (P2) where it generates between phrases, and a sentence (P3), it is classified further, and it includes in the sound of each part which constitutes the generating sound which should be specified, respectively.

[0042] sn(s) of each are the above-mentioned decryption auxiliary information s1, s2, and s3, --, a data stream prepared for every sampling period of the coding voice data a1, s2, s3, --, sn, and are a total of 1-bit data [about 4-bit] as a triplet and emphasis position identification information as information about the sound of each part which constitutes the above-mentioned generating sound. Moreover, since there is frequency which is hard to catch like the 3rd characteristic frequency region in the race in the non-English area, the specification information on the frequency band (especially center frequency) which should be emphasized may be individually included in this decryption

auxiliary information.
[0043] The synthetic section 260 adds the decryption auxiliary information s1, s2, s3, --, sn specified as the coding voice data a1, a2, a3, --, an generated by the coding section 210 as mentioned above by the analysis section 250, and writes each in a record medium 20. In addition, the voice data newly generated by the synthetic section 260 can have the various logical structures as shown in drawing 5 (a) - drawing 5 (c). For example, as shown in drawing 5 (a), the generated voice data may be the structure where the decryption auxiliary information s1, s2, s3, --, sn that it corresponded for every data of the coding voice data a1, a2, a3, --, an was added. Moreover, as shown in drawing 5 (b), the generated voice data may be structure dealt with as data of a group with which the coding voice data a1, a2, a3, --, an differs from the decryption auxiliary information s1, s2, s3, --, sn, respectively. Furthermore, as shown in drawing 5 (c), the generated voice data may be constituted by the pair of two or more groups which constitute the coding voice data a1, a2, a3, --, an, and two or more groups to which the decryption auxiliary information s1, s2, s3, --, sn corresponds.
[0044] Next, the structure of the terminal unit 400 by the side of the user who performs a decryption and reproduction of voice data equipped with the data structure concerning this invention is explained.
[0045] Drawing 6 is drawing showing the structure of the decryption section 410 of a terminal unit 400, and drawing 7 is drawing showing the structure of the PCM data generation section 415 in the coding section 410 shown in drawing 6 .
[0046] As shown in drawing 6 , voice data is incorporated by the decryption section 410 through I/O31 from external storage 30. In addition, the voice data stored in external storage 30 is distributed through information means of communications, such as a computer network and a satellite, or it is data stored in CD which the user purchased, in addition image data is also recorded in this external storage 30. Moreover, when the voice data stored in external storage 30 is compressed, data extension by software etc. is performed as pretreatment of a decryption.
[0047] In the decryption section 410, the extraction section 411 extracts the decryption auxiliary information s1, s2, s3, --, sn from the voice data read from external storage 30 first. The information (V, CF, CR, TF, TR, P1, P2, P3) about the sound of each part which constitutes generating sound among the extracted decryption auxiliary information s1, s2, s3, --, sn is inputted into the time-factor generation section 412 with the directions information from a user that it was inputted from the input means 460. Moreover, emphasis position identification information is inputted into the amplitude emphasis coefficient generation section 412 with a user's directions information that it was inputted from the input-process means 460, among the extracted decryption auxiliary information s12, s2, s3, --, sn. Furthermore, the information about the frequency component (center CH) which should be emphasized among the extracted decryption auxiliary information s12, s2, s3, --, sn is inputted into the emphasis band data generation section 414 with a user's directions information that it was inputted from the input-process means 460.
[0048] Moreover, with this operation gestalt, as a user's reproduction-speed directions information that it is inputted from the input means 460, as shown in the table of drawing 8 , two or more regeneration levels H3-S6 are prepared. By this operation gestalt, regeneration-level N is made into a standard reproduction speed (natural speed), and it is directed for the ratio of the reproduction time on the basis of this natural speed, and the scale factor of a reproduction speed so that it goes to H3 and a reproduction speed is early gone to S6 conversely, and a reproduction speed may be made late so that the front shell of drawing 8 may also be understood.
[0049] The reproduction-speed scale factor determined by the relation between a regeneration level (a user directs) as shown in drawing 9 , and the kind of sound of each part which constitutes generating sound is equipped with the table set up beforehand, and the above-mentioned time-factor generation section 411 outputs a reproduction-speed scale factor to the PCM data generation section 415 based on this table.
[0050] The above-mentioned amplitude emphasis coefficient generation section 412 is equipped with two kinds of tables as shown in drawing 10 . Drawing 10 (a) is a table applied (when there are no emphasis directions), when emphasis position identification information is not contained in the decryption auxiliary information s1, s2, s3, --, sn extracted by the extraction section 411, and drawing 10 (b) is a table applied (when there are no emphasis directions), when emphasis position

identification information is contained in the decryption auxiliary information s1, s2, s3, --, sn. In addition, the parameter shown in these tables means the scale factor on the basis of the amplitude of each frequency component of the separated coding voice data as decryption auxiliary information in the extraction section 411.

[0051] When the directions information on the frequency band (it specifies at the Center CH) which should be emphasized to the decryption auxiliary information s1, s2, s3, --, sn is included, the above-mentioned emphasis band data generation section 414 generates the parameter which changes the amplitude of each frequency component about a total of 11 CH(s) of 5CH a 5CH and high-frequency component side the low frequency component side which adjoins Center CH, as shown in drawing 11 (a). In addition, the emphasis band data generation section 414 is equipped with the table where the amplitude scale factor of the center CH according to the regeneration level was beforehand set up as shown in drawing 11 (b), and the amplitude scale factor of Center CH is determined according to the reproduction-speed directions information that it was inputted from the input means 460. Moreover, each amplitude scale factor of CH which adjoins Center CH is set up, respectively so that straight-line approximation can be carried out like drawing 11 (a) on the basis of the amplitude scale factor of this center CH, and it is outputted to the PCM data generation section 415.

[0052] The PCM data generation section 415 is equipped with the sinusoidal generator 422 made to generate the frequency component equivalent to each channel as shown in drawing 7 . From the coding voice data from the extraction section 411, a control section 421 newly generates an amplitude coefficient based on the amplitude scale-factor data from the amplitude information on each frequency component, and the emphasis band data generation section 414, and carries out the multiplication of this generated amplitude coefficient to data (a criteria amplitude is shown) from the sinusoidal generator 422 in a multiplier 423. And the PCM data decrypted when the data of each obtained frequency component made it add with an adder 424 are obtained. Furthermore, a control section 421 performs slowing and shortening of voice data which are decrypted based on the reproduction-speed scale-factor data from the time-factor generation section 412 by adjusting the number of times of an output of each of this coding voice data a1, a2, a3, --, an. Since the number of times of an output of the picture synchronizing signal D outputted for every coding voice data a1, a2, a3, --, an will also be simultaneously adjusted at this time, the display timing control of a picture becomes possible at the reproduction side of voice data.

[0053] The data decrypted in the PCM data generation section 415 as mentioned above turn into decryption data adjusted along with the time-axis according to a user's reproduction-speed directions information. In the scale-factor parameter and multiplier 416 which the front shell amplitude emphasis coefficient generation section 412 of either drawing 10 (a) or drawing 10 (b) determined, the multiplication of the data decrypted in this PCM data generation section 415 is carried out. Thereby, the voice data for reproduction is obtained. The obtained voice data for reproduction is changed into analog data by D/A converter 417, and is outputted from a loudspeaker 480 as voice of the reproduction speed which the user directed.

[0054] The display of the image data read from external storage 30 with the terminal unit 400 on the other hand is also possible. Drawing 12 is drawing showing the structure of a bit mapped display.

[0055] The bit mapped display is equipped with the memory 432 (VRAM) which stores 1 or the frame beyond it, and the drawing section 431 writes the image data (when compressed, data extension of the software etc. is carried out by 32) read from external storage 30 through I/O32 in these memory 432. The image data written in memory 432 is displayed on a display 470 through switch S/W433 for every frame. In addition, write-in timing of these drawing section 431 and change timing of S/W433 are performed by the timing controller 434.

[0056] Timing of voice reproduction and image display is performed with this operation gestalt by counting the picture synchronizing signal D outputted from the PCM data generation section 415 as shown in drawing 13 (a). That is, when the PCM data generation section 415 will generate late voice **** for reproduction of a reproduction speed as shown in drawing 13 (b) if data rewriting of memory 432 will be performed every the case of voice reproduction by natural speed, 3 [for example,], clock, data rewriting suitable for the delay timing of this voice data is attained (it becomes possible to make the display timing of a picture in agreement with the timing of voice reproduction).

[0057] That is, it is one frame or its N times of two or more image data stored in memory 432 with this operation form (it is a positive rational number). the criteria reproduction time period determined based on Tv and the audio digitization sampling period (for example, sound clock of Music CD) in the criteria rewriting period of the image data of N being one half and 2/3 -- Ta -- and When making the reproduction time period determined for every sound of each part which constitutes generating sound based on the directed reproduction-speed information into Ta' (>Ta), It is characterized by stopping next image data rewriting operation from the end point in time of image data rewriting operation to predetermined timing to Tvx (Ta'/Ta) (-1).

[0058] In addition, timing adjustment with voice reproduction and image display is not limited to an above-mentioned operation gestalt. For example, when making the reproduction time period determined for every sound of Ta and each part which constitutes generating sound based on the directed reproduction-speed information in the criteria reproduction time period determined based on the audio digitization sampling period into Ta' (>Ta), it rewrites and the average rewriting frequency of the image data stored in the aforementioned memory may be set up the twice which was specified beforehand and which is frequency (Ta'/Ta).

[0059]

[Effect of the Invention] As mentioned above, according to this invention, it is incorporated from a microphone etc. or voice data is constituted by decryption auxiliary information including the kind of sound of each part which constitutes the generating sound referred to in the case of a decryption of the coding voice data encoded according to the predetermined rule from the sound signal accumulated in the past, and this coding voice data etc. By providing a user with such voice data by a predetermined record medium and the predetermined distribution method, two or more kinds of voice data for reproduction from which the speed set up arbitrarily differs can be decrypted by the user side. The amount of data of the voice data which should be offered from a data provider to a user can be reduced by this, and saving of the amount of records of a record medium and shortening of data distribution time are realized.

[0060] Furthermore, the image reconstruction suitable for the reproduction speed of the decrypted voice data for reproduction also becomes possible by adding a picture synchronizing signal to the above-mentioned coding voice data with decryption auxiliary information.

[Translation done.]

* NOTICES *

1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]
[Drawing 1] It is drawing for explaining each operation gestalt of this invention notionally.
[Drawing 2] (a) is the block diagram showing the outline composition of the voice data generation equipment which realizes the generation method of the voice data concerning this invention, and (b) is the block diagram showing the outline composition of the voice data regenerative apparatus which realizes the reproduction method of the voice data concerning this invention.
[Drawing 3] It is the block diagram showing the composition of the coding section in the voice data generation equipment shown in drawing 2 (a).
[Drawing 4] It is drawing for explaining notionally a part of decryption auxiliary information required for a decryption of the encoded voice data.
[Drawing 5] It is drawing for explaining notionally the data structure of the voice data concerning this invention.
[Drawing 6] It is the block diagram showing the composition of the voice data regenerative apparatus (terminal unit) which realizes the reproduction method of the voice data concerning this invention.
[Drawing 7] It is the block diagram showing the composition of the PCM data generation section in the voice data regenerative apparatus shown in drawing 6 .
[Drawing 8] It is the table which was set up for every regeneration level and in which showing an example of the ratio of the reproduction time on the basis of the regeneration level of natural speed, and the scale factor of a reproduction speed.
[Drawing 9] It is the table referred to in the time-factor generation section shown in drawing 6 , and is the table having shown an example of the reproduction speed set up for every kind of sound of each part which constitutes generating sound for the scale factor on the basis of the regeneration level of natural speed.
[Drawing 10] It is the table referred to in the amplitude emphasis coefficient generation section shown in drawing 6 , and is the table having shown an example of the amplitude set up for every kind of sound of each part which constitutes generating sound for the scale factor on the basis of the regeneration level of natural speed.
[Drawing 11] (a) is a table referred to in the emphasis band data generation section shown in drawing 6 , it is drawing for explaining edit operation of the directed frequency band data, and (b) is the table having shown the amplitude of the directed frequency band (center CH) for the scale factor on the basis of the regeneration level of natural speed.
[Drawing 12] It is drawing showing the composition of the display which displays image data synchronizing with the voice reproduction by the voice data regenerative apparatus (terminal unit) shown in drawing 6 .
[Drawing 13] It is a timing diagram for explaining the image display timing which synchronized with voice reproduction operation.
[Description of Notations]
10 [-- Voice data generation equipment, 210 / -- The coding section, 250 / -- The analysis section, 260 / -- The synthetic section, 400 / -- A terminal unit (PC), 410 / -- Decryption section.] -- 20 The information source, 30,310 -- A record medium, 200

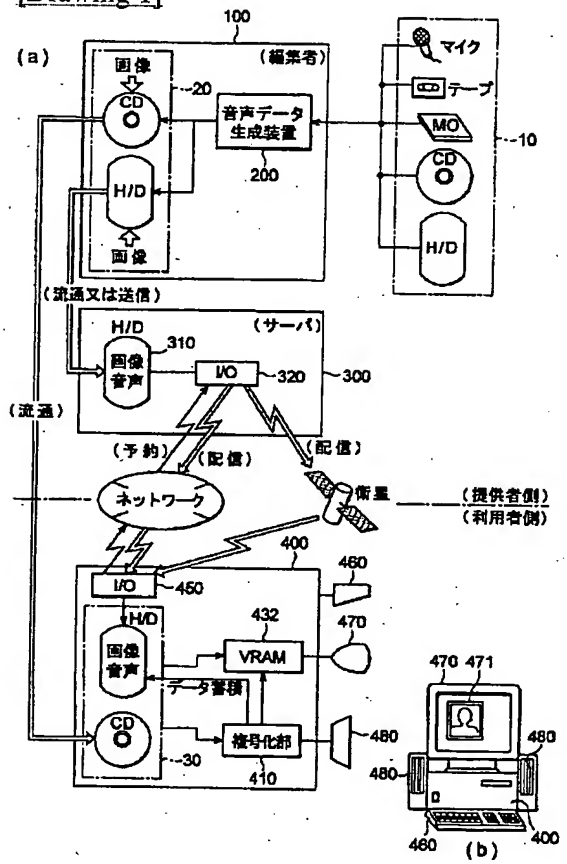
[Translation done.]

* NOTICES *

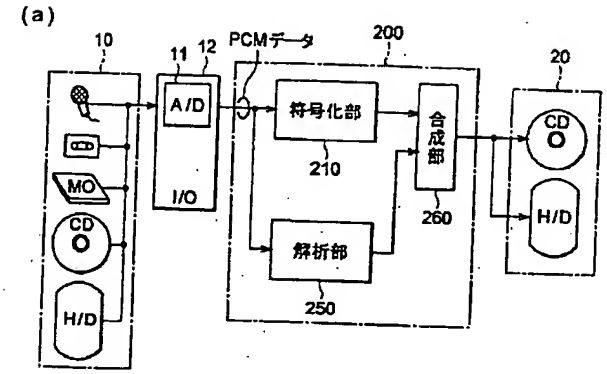
1.This document has been translated by computer. So the translation may not reflect the original precisely.
2.**** shows the word which can not be translated.
3.In the drawings, any words are not translated.

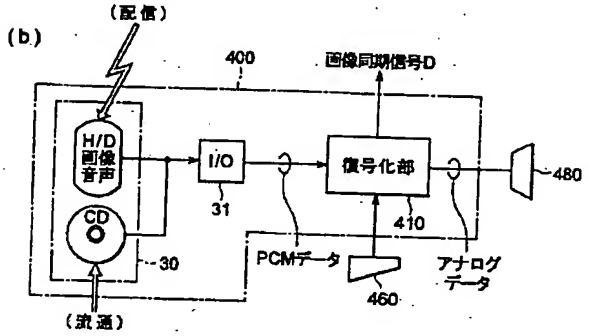
DRAWINGS

[Drawing 1]

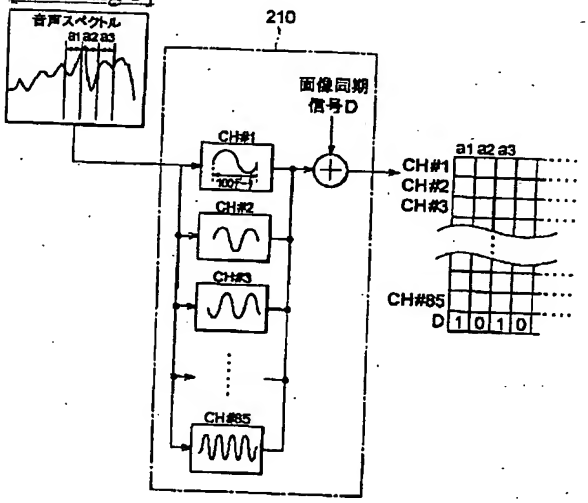


[Drawing 2]

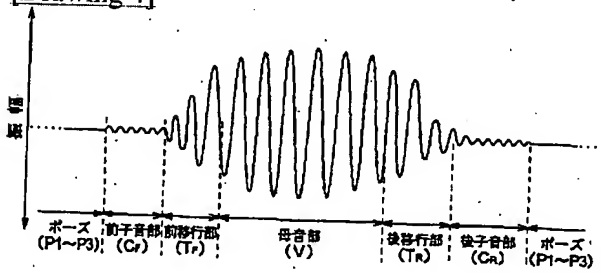




[Drawing 3]



[Drawing 4]

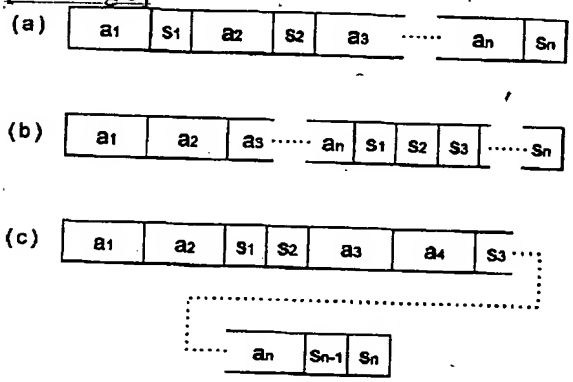


[Drawing 8]

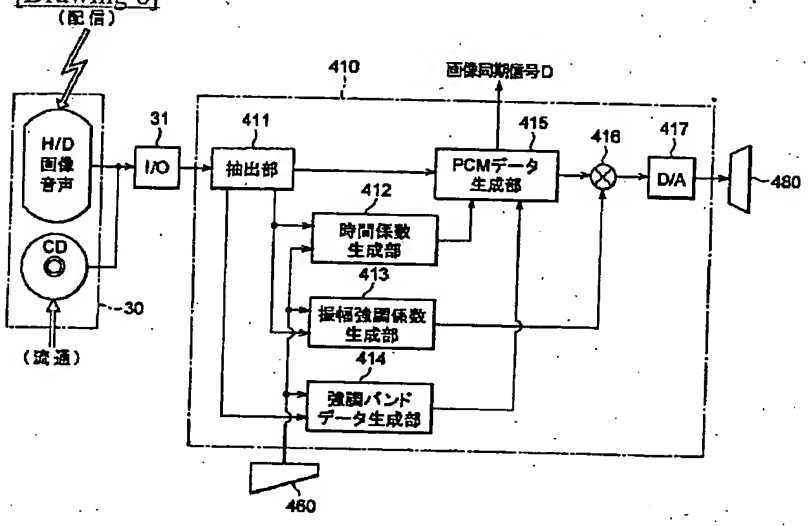
再生レベル	H3	H2	N	S2	S3	S4	S5	S6
再生時間比	0.64	0.80	1.00	1.25	1.56	1.95	2.44	3.05
再生速度倍率	1.56	1.25	1.00	0.80	0.64	0.51	0.41	0.33

速い← →遅い

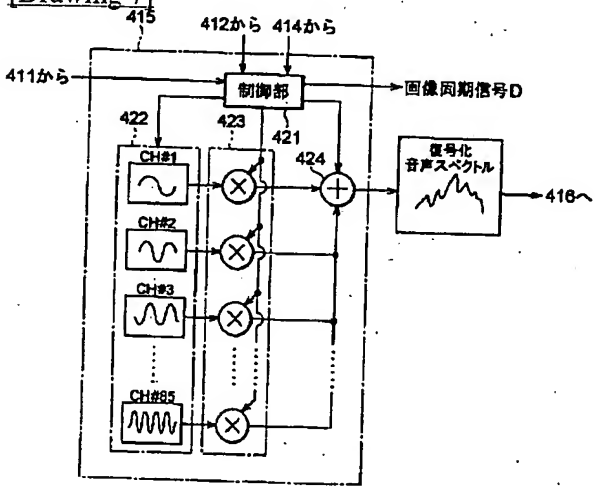
[Drawing 5]



[Drawing 6]



[Drawing 7]

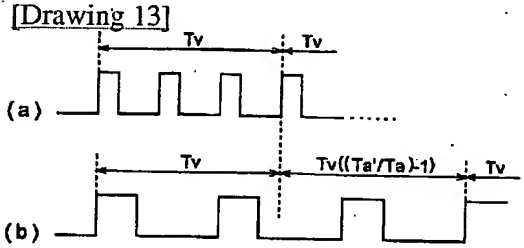


[Drawing 9]

再生速度倍率 ＼ 発生音	CF	TF	V	TR	CR	P1	P2	P3
0.33	0.64	0.64	0.25	0.64	0.64	0.33	0.80	0.80
0.41	0.64	0.64	0.33	0.64	0.64	0.41	0.80	0.80
0.51	0.64	0.64	0.41	0.64	0.64	0.51	0.80	0.80
0.64	0.64	0.64	0.51	0.64	0.64	0.64	0.80	0.80
0.80	0.80	0.80	0.64	0.80	0.80	0.80	0.80	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.25	1.10	1.10	1.30	1.10	1.10	1.10	1.05	1.00
1.56	1.25	1.25	1.70	1.25	1.25	1.25	1.10	1.00

遅い ↑ / ↓ 速い

[Drawing 13]



[Drawing 10]

(a)

再生速度倍率 ＼ 発生音	CF	TF	V	TR	CR
0.33	2.00	1.50	1.00	1.50	2.00
0.41	1.75	1.32	1.00	1.32	1.75
0.51	1.52	1.23	1.00	1.23	1.52
0.64	1.32	1.15	1.00	1.15	1.32
0.80	1.15	1.07	1.00	1.07	1.15
1.00	1.00	1.00	1.00	1.00	1.00
1.25	1.15	1.10	1.00	1.10	1.15
1.56	1.15	1.10	1.00	1.10	1.15

遅い ↑ / ↓ 速い

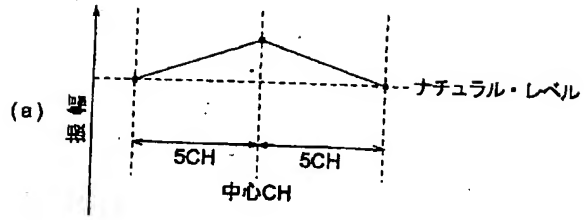
(b)

再生速度倍率 ＼ 発生音	CF	TF	V	TR	CR
0.33	2.40	1.95	1.30	1.95	2.40
0.41	2.28	1.72	1.30	1.72	2.28
0.51	2.00	1.60	1.30	1.60	2.00
0.64	1.72	1.50	1.30	1.50	1.72
0.80	1.31	1.22	1.14	1.22	1.31
1.00	1.00	1.00	1.00	1.00	1.00
1.25	1.31	1.22	1.14	1.22	1.31
1.56	1.31	1.22	1.14	1.22	1.31

遅い ↑ / ↓ 速い

[Drawing 11]

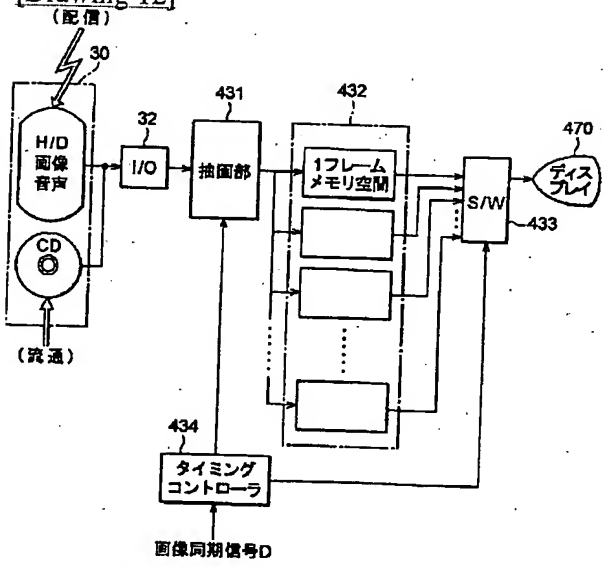
(a)



(b)

再生速度倍率	中心CH
0.33	1.20
0.41	1.20
0.51	1.20
0.64	1.12
0.80	1.06
1.00	1.00
1.25	1.00
1.56	1.00

[Drawing 12]



[Translation done.]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号
特開2001－242890
（Ｐ２００１－２４２８９０Ａ）
(43)公開日　平成13年９月７日(2001.9.7)

(51)Int.Cl.[7]	識別記号	ＦＩ		テーマコード（参考）
G10L 19/00		G09B 5/04		2C028
G09B 5/04		G11B 20/02	G	5C053
G11B 20/02		20/10	301 Z	5C064
20/10	301	H04N 7/173	630	5D044
H04N 5/92		G10L 9/18	A	5D045

審査請求　未請求　請求項の数12　ＯＬ　（全13頁）　最終頁に続く

(21)出願番号　特願2000－51801（Ｐ2000－51801）

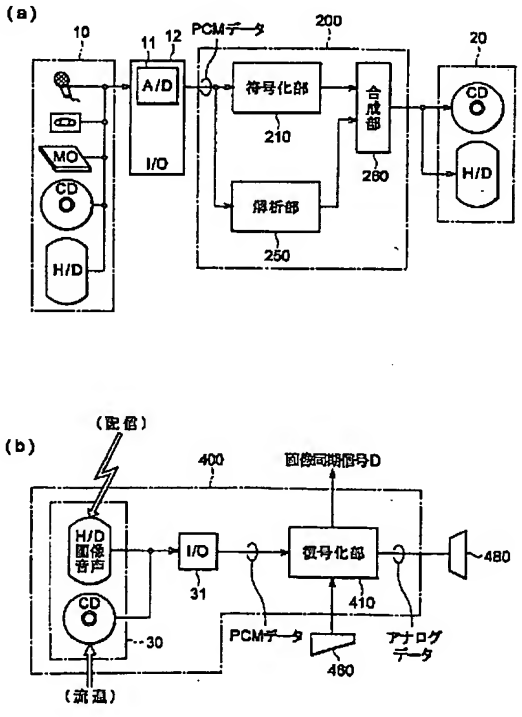
(22)出願日　平成12年２月28日(2000.2.28)

(71)出願人　000104179
カナース・データー株式会社
東京都千代田区東神田１丁目10番７号
(72)発明者　関口　博司
東京都中野区松が丘２－28－１
(74)代理人　100088155
弁理士　長谷川　芳樹　（外２名）

最終頁に続く

(54)【発明の名称】音声データのデータ構造、生成方法、再生方法、記録方法、記録媒体、配信方法、及びマルチメディアの再生方法

(57)【要約】
【課題】　聞き取り易さを損なうことなく、利用者側で再生速度が変更された音声データの復号化を可能にする音声データのデータ構造等を提供する。
【解決手段】　合成部(260)が、符号化部(210)が音声信号から所定の規則に従って符号化した符号化音声データに、解析部(250)により特定された該符号化音声データの復号化の際に参照される発生音を構成する各部の音の種類等を含む復号化補助情報を付加する。これにより、利用者側にて任意の再生速度の音声データの復号化が可能となり、情報通信技術を利用したデータ配信サービス等のコンテンツとして有望な音声データとなり得る。



1

【特許請求の範囲】

【請求項１】　音声信号から所定の規則に従って符号化された符号化音声データと、

前記符号化音声データの復号化の際に参照される情報であって、前記音声信号の波動に関する物理量から特定された、少なくとも発生音を構成する各部の音の種類に関する情報を含む復号化補助情報とを備えた音声データのデータ構造。

【請求項２】　前記音声信号の波動に関する物理量には、前記音声信号の周波数スペクトル情報が含まれることを特徴とする請求項１記載の音声データのデータ構造。

【請求項３】　前記復号化補助情報には、振幅方向に強調すべき時間軸方向の位置を指示するための強調位置識別情報が含まれることを特徴とする請求項１又は２記載の音声データのデータ構造。

【請求項４】　前記復号化補助情報には、前記復号化された音声データの周波数成分のうち、強調すべき周波数成分を個別に指示する情報が含まれることを特徴とする請求項１～３のいずれか一項記載の音声データのデータ構造。

【請求項５】　所定の表示手段に表示されるべき画像データの表示タイミングを指示する情報を含むことを特徴とする請求項１～４のいずれか一項記載の音声データのデータ構造。

【請求項６】　音声信号から所定の規則に従って符号化された符号化音声データを生成する第１行程と、

前記符号化音声データの復号化の際に参照される復号化補助情報として、前記音声信号の波動に関する物理量から少なくとも発生音を構成する各部の音の種類に関する情報を特定する第２行程と、

前記第１行程において生成された前記符号化音声データに前記第２行程において特定された前記復号化補助情報を付加することにより、前記音声データとして新たな合成データを生成する第３工程とを備えた音声データの生成方法。

【請求項７】　請求項１記載のデータ構造を有する音声データを、予め指定された再生速度情報に基づいて発生音を構成する各部の音ごとに調節された速度で音声を再生する音声データの再生方法であって、

前記音声データから、前記符号化音声データの復号化の際に参照される復号化補助情報を抽出する第１行程と、

前記第１行程において抽出された復号化補助情報に含まれる少なくとも発生音を構成する各部の音に関する情報を参照しながら、前記符号化音声データに含まれる該発生音を構成する各部の音ごとに音声再生に適した再生速度を前記予め指定された再生速度情報を基準にして決定する第２行程と、

前記符号化音声データに含まれる発生音を構成する各部の音ごとに、前記決定された再生速度に相当するよう該

2

符号化音声データの該当部分に対して伸長処理又は短縮処理を施しながら、該符号化音声データを復号化する第３行程とを備えた音声データの再生方法。

【請求項８】　音声信号から所定の規則に従って符号化された符号化音声データと、

前記符号化音声データの復号化の際に参照される情報であって、前記音声信号の波動に関する物理量から特定された、少なくとも発生音を構成する各部の音の種類に関する情報を含む復号化補助情報とを、所定の記録媒体に記録する音声データの記録方法。

【請求項９】　音声信号から所定の規則に従って符号化された符号化音声データと、

前記符号化音声データの復号化の際に参照される情報であって、前記音声信号の波動に関する物理量から特定された、少なくとも発生音を構成する各部の音の種類に関する情報を含む復号化補助情報とが記録された音声データの記録媒体。

【請求項１０】　音声信号から所定の規則に従って符号化された符号化音声データと、

前記符号化音声データの復号化の際に参照される情報であって、前記音声信号の波動に関する物理量から特定された、少なくとも発生音を構成する各部の音の種類に関する情報を含む復号化補助情報とを、有線又は無線の情報伝達手段を介して通信相手に送信する音声データの配信方法。

【請求項１１】　１又はそれ以上の画像データを一旦メモリ上に展開し、該メモリ上に格納された画像データのうち１フレーム分の画像データを、請求項１記載のデータ構造を有する音声データの再生動作に同期して、順次所定の表示手段に表示するマルチメディアの再生方法であって、

前記メモリに格納される前記複数の画像データのうち１フレーム分又はそのＮ倍（正の有理数）の画像データの基準書き換え周期をＴｖ、音声のディジタル化サンプリング周期に基づいて決定された基準再生時間周期をＴａ、そして、指示された再生速度情報に基づいて発生音を構成する各部の音ごとに決定された再生時間周期をＴａ´（＞Ｔａ）とするとき、

所定タイミングでの画像データ書き換え動作の終了時点からＴｖ×（（Ｔａ´／Ｔａ）－１）まで、次回の画像データ書き換え動作を休止させるマルチメディアの再生方法。

【請求項１２】　１又はそれ以上の画像データを一旦メモリ上に展開し、該メモリ上に格納された画像データのうち１フレーム分の画像データを、請求項１記載の音声データの記録媒体に記録されている音声データの再生動作に同期して、順次表示手段に表示するマルチメディアの再生方法であって、

音声のディジタル化サンプリング周期に基づいて決定された基準再生時間周期をＴａ、そして、指示された再生

速度情報に基づいて発生音を構成する各部の音ごとに決定された再生時間周期をＴａ´（＞Ｔａ）とするとき、前記メモリに格納される画像データの平均書き換え周波数を、予め指定された書き換え周波数の（Ｔａ´／Ｔａ）倍に設定するマルチメディアの再生方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】この発明は、聞き取り易さを損なうことなく、利用者側で任意に再生速度が変更された再生用音声データの復号化を可能にする音声データのデータ構造、生成方法、再生方法、記録方法、記録媒体、配信方法、及び該音声データの再生速度に同期したマルチメディアの再生方法に関するものである。

【０００２】

【従来の技術】従来から、英会話等の語学の独習用、詩吟の練習用、法律の独習用、歌の練習、その他の目的のために、カセットテープ等の記録媒体に音楽とともに音声情報が記録された教材が種々提供されている。ここで、英会話の独習用の教材を例に説明すると、従来の主な記録媒体は、例えば一連の英語の発声（音声情報）が記録されてたカセットテープ（又はＣＤ）であり、学習者はこのテープ教材とテキストとを組み合せて使用していた。なお、このような教材には、初級用から上級用まで種々のレベルが用意されている。

【０００３】また、日本国特許第２５８１７００号には、複数の区画に区分された上級者学習用に適した音声情報列（ナチュラル・スピードの発生音を構成する各部の音）が記録された第１領域と、これら各区画に対応した等価な区画からなる初級者学習用に適した音声情報列（はっきりとした発生音を構成する各部の音であって、言語学上は同一の意味で派生の異なる音声）が記録された第２領域と、該上級者学習用及び初級者学習用の各音声情報列の対応する各区画の関係を、これら音声情報列の各区画の記録媒体における記録位置で示す情報が記録された第３領域とを、少なくとも備えたＣＤ－ＲＯＭ等の情報記録媒体、及びこのような構造を備えた情報記録媒体の対応する区画間での切替え再生等を含む再生方法が提案されている。

【０００４】

【発明が解決しようとする課題】上述のように、日本国特許第２５８１７００号の情報記録媒体には、該媒体上の第１領域にネイティブスピーカーの発生音を構成する各部の音が記録され、また第２領域に言語上は同一の意味で遅緩した発音で構成された音声情報列が記録されている。したがって、第１領域に記録された音声情報列が再生されている最中に再生音を聞き取れなかった場合、第２領域に記録された同一内容の音声情報列（第１音声情報列の再生中の区画と第２音声情報列の再生すべき区画との対応は第３領域に記録されている）を切替えて再生することにより、学習者は聞き取れなかった音声を認知することができる。また、近年のパーソナル・コンピュータ等の情報処理機器の普及・高性能化を考慮すれば、制作時間の短縮や制作コストの削減のため、上記第１領域に記録される音声情報列から遅緩した第２領域に記録される音声情報列を生成することも不可能ではない。

【０００５】しかしながら、単にネイティブ・スピーカーの音声を時間軸に沿って均一に伸張させたのでは、利用者の聞き取り易さを損なってしまう。すなわち、主として日本人がナチュラル・スピードの英語を単にゆっくり再生して聴けるようにした場合であっても、各周波数成分について単純にかつ一様に音声再生時間を伸ばしたり短縮したのでは不充分であり、発生音を構成する各部の音、例えば子音部のスペクトルの時間変化が言語上の音として別の音を意味する可能性があるからである。例えば、ＢＡ（バ）とＰＡ（パ）の発音は、前者のスペクトル変化が速く、後者は遅いだけでスペクトルそのものはほとんど同じ形をしており、ＢＡ（バ）という発音の子音部も含めて時間を伸長するとＰＡ（パ）と聴こえることになる。

【０００６】一方、学習者のヒヤリング・レベルも、例えば上記第２領域に記録された音声再生速度でも十分に聞き取れない者、提供されたナチュラル・スピードでは満足できない者など様々であり、このように異なるヒヤリング・レベルの学習者を個々に満足させようとすると、各学習者のヒヤリング・レベルに応じた複数種類の音声情報を予め用意しなければならない。しかしながら、現状では学習者側で自分のヒヤリング・レベルに合った音声情報を選択できず、また、各学習者のヒヤリング・レベルに合った複数種類の音声情報を用意することは、ＣＤ等の記録媒体の記録容量に限界があるため現実的ではない。

【０００７】さらに、近年の情報通信技術の発達により、インターネット等のコンピュータ・ネットワークを利用したデータ配信も注目されている。このようなデータ配信を利用した音声情報の提供を考える場合、大量のデータを送信するにはまだまだ通信時間や通信コストの面で実用レベルに達しているとは言えない。

【０００８】この発明は上述のような課題を解決するためになされたもので、聞き取り易さを損なうことなく、利用者側で希望する再生速度の再生用音声データの復号化を可能にする音声データのデータ構造、生成方法、該音声データから利用者側で希望する再生スピードの新たな再生用音声データを復号化するための音声データの再生方法、該音声データを所定の記録媒体に記録するための音声データの記録方法、該音声データが記録された記録媒体、該音声データをコンピュータ・ネットワークや衛生通信システムを利用して利用者に提供するための音声データの配信方法、及び該音声データの再生動作に同期した画像表示を可能にするためのマルチメディアの再

生方法を提供することを目的としている。

【０００９】

【課題を解決するための手段】上述の課題を解決するため、この発明に係る音声データのデータ構造は、音声信号から所定の規則に従って符号化された符号化音声データと、該符号化音声データの復号化の際に参照される復号化補助情報とを備える。

【００１０】特に、上記復号化補助情報は、音声信号の波動に関する物理量、例えば周波数スペクトル情報などから特定された、少なくとも発生音を構成する各部の音の種類に関する情報を含む。これは、上述のような子音部の変化により異なる音に聞こえてしまう不具合を解消するためである。子音部の伸長度をＢＡ（バ）と聴こえる限界に留め、母音部のみ望みの音声再生時間に伸長あるいは短縮するようにすれば、ＢＡ（バ）のままに聴こえることになる。母音部はいくら伸長あるいは短縮してもその母音のままで聴こえるから望みの長さ（望みの再生時間）に設定できる。

【００１１】一方、非英語圏の国民には弱すぎて聴き取りにくい音声や特定の周波数成分だけを選択的に２倍とか３倍に強調して聴かせることも語学学習等には必要である。母音部も含めて強調したのでは全体が大きくなり過ぎて効果がない。どうしても選択的に強調しなければならない。そこで、上記復号化補助情報は、強調すべき位置を指示するための強調位置識別情報を含むのが好ましい。また、この復号化補助情報は、復号化された音声データの周波数成分のうち、強調すべき周波数成分を個別に指示する情報を含むようにしてもよい。

【００１２】なお、上記符号化音声データを得るための符号化としては、異なる再生速度でも聞き取り易くするため、予め符号化対象の音声信号を周波数成分に分解し、該分割された周波数成分ごとにその振幅情報等をデータ化する特願平１０－２４９６７２号記載の符号化が適している。また、符号化対象である音声信号は、ディジタル化された電気信号であり、その情報源としては、マイクを介して取り込まれたアナログ音声情報、磁気テープ等から読み出されたアナログ音声情報、ＭＯ、ＣＤ、ハードディスク等に記録されたディジタル音声情報のいずれであってもよい。ただし、アナログ音声情報の場合は、一旦Ａ／Ｄ変換される必要がある。また、ＣＤ等に記録されたデータが圧縮符号化されている場合には、該圧縮データを伸張（解凍）する必要がある。

【００１３】近年普及し始めたインターネット等のコンピュータネットワーク、ケーブルＴＶネットワーク、衛生通信などの分野に着目すると、文字データ、音声データ、静止画データ、動画データなどのマルチメディアによる情報提供サービスも広く行われるようになってきており、このような情報通信技術を利用した情報提供サービスの１つとして、この発明を適用させるためには、画像データの表示タイミングの調節が不可欠となる。そこで、上述のようなデータ構造を備えた音声データに、所定の表示手段に表示されるべき画像データの表示タイミングを指示する情報を含めることにより、音声データの再生動作に同期させた画像表示（特に、動画表示）が可能になる。

【００１４】また、この発明に係る音声データの生成方法は、上記符号化音声データを生成する第１行程と、上記復号化補助情報を特定する第２行程と、該符号化音声データに復号化補助情報を付加する第３行程とを備える。上記第１行程では、例えば特願平１０－２４９６７２号に記載された符号化技術のように、所定の規則に従って音声信号の符号化が行われる。上記第２行程では、符号化音声データの復号化の際に参照される復号化補助情報として、音声信号の波動に関する物理量（例えば周波数スペクトル情報）から少なくとも発生音を構成する各部の音の種類に関する情報が特定される。なお、上記第１及び第２行程は並行して実施することも可能である。

【００１５】以上のように生成された音声データ（符号化音声データを含む）を利用することにより、該音声データの復号化において、指定された再生速度情報に基づいた発生音を構成する各部の音ごとに調節された各学習者のヒヤリング・レベルに合った再生用音声データの復号化が可能になる。このように復号化された再生用音声データを再生することにより、学習者は自己の指定した速度に調節された再生音声を聴くことができる。すなわち、この発明に係る音声データの再生方法は、上述のように生成された音声データから復号化補助情報を抽出する第１行程と、抽出された抽出された復号化補助情報に含まれる少なくとも発生音を構成する各部の音に関する情報を参照しながら、発生音を構成する各部の音ごとに再生速度を決定する第２行程と、決定された再生速度に相当するよう符号化音声データの該当部分に対して伸張処理（弛緩した音声の再生のため）短縮処理（より早い音声の再生のため）を施しながら、該符号化音声データを復号化する第３行程とを備える。

【００１６】なお、当該再生方法は、聞き取り易さを損なうことなく、再生速度の異なる複数種類の再生用音声データが用意できるため、生成された再生速度の異なる再生用音声データ間での切り替え再生を行いながらの学習も可能になる。また、上記発生音を構成する各部の音には、音声スペクトルにおける母音部、該母音部の前後に現れる子音部、該母音部と子音部との間に現れる移行部、音声の切れ目（ポーズ）などが含まれる。

【００１７】加えて、上述のようなデータ構造を備えた音声データの提供は、一旦ＣＤ等の記録媒体に記録された形態で利用者に提供される場合と、情報通信手段を介して利用者に提供される場合が考えられる。情報通信技術を利用する場合でも音声データの取り扱いはハードディスク等への一時記録が不可欠であり、この発明に係る

音声データの記録方法では、聞き取り易さを損なうことなく、利用者側で再生速度が変更された再生用音声データの復号化が可能になるよう、符号化音声データとともに復号化補助情報が所定の記録媒体に記録される。なお、以上の記録方法により得られる音声データの記録媒体において、符号化音声データが記録される領域と復号化補助情報が記録される領域は、異なっていてもよい。

【００１８】上述のようなデータ構造を備えた音声データを有線又は無線の情報伝達手段を介して通信相手に送信するデータ配信方法として、この発明に係る音声データの配信方法では、音声信号から所定の規則に従って符号化された符号化音声データと、該符号化音声データの復号化の際に参照される情報であって、音声信号の波動に関する物理量から少なくとも発生音を構成する各部の音の種類に関する情報を含む復号化補助情報とを、通信相手に送信する。なお、この発明に係る音声データの配信方法では、送信される符号化音声データと復号化補助情報とを別個に通信相手に送信する構成であってもよい。

【００１９】加えて、上述のようなデータ構造を備えた音声データの再生動作を文字データ、画像データ（静止画と動画）などのマルチメディア再生に対応させる場合、特に動画再生と、再生速度が自由に変更された音声の再生との同期をとることが重要である。すなわち、動画は例えば１秒間に２０フレーム程度の画像データをディスプレイに順次表示していくが、音声の再生動作との表示タイミングとの同期がとれていないと不自然な表示動作になってしまう。

【００２０】そこで、この発明に係るマルチメディアの再生方法は、１又はそれ以上の画像データを一旦メモリ上に展開し、該メモリ上に格納された画像データのうち１フレーム分の画像データを、上述のようなデータ構造を備えた音声データの再生動作に同期して、順次所定の表示手段に表示することを特徴としている。

【００２１】具体的に、当該マルチメディアの再生方法は、メモリに格納される複数の画像データのうち１フレーム分又はそのＮ倍（正の有理数）の画像データの基準書き換え周期をＴｖ、音声のディジタル化サンプリング周期に基づいて決定された基準再生時間周期をＴａ、そして、指示された再生速度情報に基づいて発生音を構成する各部の音ごとに決定された再生時間周期をＴａ´（＞Ｔａ）とするとき、所定タイミングでの画像データ書き換え動作の終了時点からＴｖ×（（Ｔａ´／Ｔａ）－１）まで、次回の画像データ書き換え動作を休止させる。

【００２２】なお、画像データの表示タイミングの調節は、音声のディジタル化サンプリング周期に基づいて決定された基準再生時間周期をＴａ、そして、指示された再生速度情報に基づいて発生音を構成する各部の音ごとに決定された再生時間周期をＴａ´（＞Ｔａ）とすると、メモリに格納される画像データの平均書き換え周波数を、予め指定された書き換え周波数の（Ｔａ´／Ｔａ）倍に設定するようにしても、音声データの再生動作に同期した画像表示が可能になる。

【００２３】

【発明の実施の形態】以下、この発明に係る音声データのデータ構造等の各実施形態を図１～図１３を用いて説明する。なお、図面の説明において同一部分には同一符号を付して重複する説明は省略する。

【００２４】この発明に係るデータ構造を備えた音声データは、再生時の聞き取り易さを損なうことなく、利用者が自由に設定した再生速度の新たな再生用音声データの復号化を、該利用者側で行うことを可能にする。このような音声データの利用形態は、近年のディジタル技術の発達やデータ通信環境の整備により種々の態様が考えられる。図１は、この発明に係るデータ構造を備えた音声データがどのように産業上利用されるかを説明するための概念図である。

【００２５】図１（ａ）に示されたように、この発明に係るデータ構造を備えた音声データの生成に利用される情報源１０としては、例えばマイクを介して直接取り込まれたり、既に磁気テープなどに記録されたアナログ音声情報、さらにはＭＯ、ＣＤ（ＤＶＤを含む）、Ｈ／Ｄ（ハードディスク）等に記録されているディジタル音声情報が利用可能であり、具体的には、市販されている教材やテレビ局、ラジオ局などから提供される音声情報などでも利用可能である。編集者１００は、このような情報源１０を利用して音声データ生成装置２００により、この発明に係るデータ構造を備えた音声データを生成する。なお、この際、現状のデータ提供方法を考えると、生成された音声データはＣＤ（ＤＶＤを含む）、Ｈ／Ｄなどの記録媒体２０に一旦記録された状態で利用者に提供される場合が多い。また、これらＣＤやＨ／Ｄには当該音声データとともに関連する画像データが記録される場合も十分に考えられる。

【００２６】上記音声データ生成装置２００により生成される音声データは、上記情報源１０から取り出されたディジタル音声信号から所定の規則に従って符号化された符号化音声データと、この符号化音声データの復号化の際に参照される情報であって、該音声信号の波動に関する物理量から特定された、少なくとも発生音を構成する各部の音の種類に関する情報を含む復号化補助情報とを備えた新たな音声データである。なお、符号化音声データを得るための符号化としては、例えば、異なる再生速度でも聞き取り易くするため、予め符号化対象の音声信号を周波数成分に分解し、該分割された周波数成分ごとにその振幅情報等をデータ化する特願平１０－２４９６７２号記載の符号化が利用可能である。生成された符号化音声データ及び復号化補助情報は、音声データ生成装置２００により、記録媒体２０に格納される。これに

より、ＣＤ、ＤＶＤ、Ｈ／Ｄ等の記録媒体には上述の符号化音声データ、復号化補助情報とともに画像データ、文字データなどのマルチメディアが記録される。

【００２７】特に、記録媒体２０としてのＣＤやＤＶＤは、雑誌の付録として利用者に提供されたり、コンピュータ・ソフト、音楽ＣＤなどと同様に店舗にて販売されるのが一般的である（市場での流通）。また、生成された音声データはサーバ３００から有線、無線を問わず、インターネット、衛生通信などの情報通信手段を介して利用者に配信される場合も十分に考えられる。

【００２８】データ配信の場合、上記音声データ生成装置２００により生成された音声データは、サーバ３００の記憶装置３１０（例えばＨ／Ｄ）に画像データなどとともに一旦蓄積される。そして、Ｈ／Ｄ３１０に一旦蓄積された音声データは、送受信装置３２０（図中のＩ／Ｏ）を介して利用者端末４００に送信される。利用者端末４００側では、送受信装置４５０を介して受信された音声データが一旦Ｈ／Ｄ（外部記憶装置３０に含まれる）に格納される。一方、ＣＤやＤＶＤ等を利用したデータ提供では、利用者が購入したＣＤを端末装置４００のＣＤドライブやＤＶＤドライブに装着することにより該端末装置の外部記録装置３０として利用される。

【００２９】通常、利用者側の端末装置４００には入力装置４６０、ＣＲＴ、液晶などのディスプレイ４７０、スピーカー４８０が装備されており、外部記憶装置３００に画像データなどとともに記録されている音声データは、当該端末装置４００の復号化部４１０（ソフトウエアによっても実現可能）によって、利用者自身が指示した再生速度の音声データに一旦復号化された後、スピーカー４８０がら出力される。一方、外部記憶装置３００に格納された画像データは一旦ＶＲＡＭ４３２に展開された後にディスプレイ４７０に各フレームごと表示される（ビットマップ・ディスプレイ）。なお、復号化部４１０により復号化された再生用音声データを上記外部記憶装置３０内に順次蓄積することにより、該外部記憶装置３０内には再生速度の異なる複数種類の再生用音声データを用意すれば、日本国特許第２５８１７００号に記載された技術を利用して再生速度の異なる複数種類の音声データ間の切り替え再生が利用者側で可能になる。

【００３０】利用者は、図１（ｂ）に示されたように、ディスプレイ４７０上に関連する画像４７１を表示させながらスピーカー４８０から出力される音声を聴くことになる。この際、音声のみ再生速度が変更されていたのでは、画像の表示タイミングがずれてしまう可能性がある。そこで、復号化部４１０が画像データの表示タイミングを制御できるよう、上記音声データ生成装置２００において生成される符号化音声データに画像表示タイミングを指示する情報を予め付加しておくのが好ましい。

【００３１】次に、図１（ａ）に示された音声データ生成装置２００及び音声データ再生装置（端末装置４００）の詳細な構造を図２を用いて説明する。なお、図２（ａ）は、音声データ生成装置２００の構成を示す図であり、図２（ｂ）は、音声データ再生装置としての端末装置４００の構成を示す図である。

【００３２】図２（ａ）に示されたように、音声データ生成装置２００に取り込まれる音声信号は情報源１０から提供される。なお、この情報源１０から提供される音声情報のうち、マイクから取り込まれる音声情報及び磁気テープからの音声情報は、ともにアナログ音声データであるため、当該音声データ生成装置２００へ入力される前にＡ／Ｄコンバータ１１（Ｉ／Ｏ１２に含まれる）によりＰＣＭデータに変換される。また、ＭＯ、ＣＤ（ＤＶＤ含む）、Ｈ／Ｄに既に格納された音声情報は、ＰＣＭデータとしてＩ／Ｏ１２を介して当該音声データ生成装置２００に取り込まれる。取り込まれた音声データが圧縮されている場合には、一旦ソフトウエア等の解凍しておく必要がある。

【００３３】音声データ生成装置２００は、上述のように前処理された情報源１０からの音声信号（電気信号）から所定の規則に従って符号化された符号化音声データを生成する符号化部２１０と、この符号化音声データの復号化の際に参照される復号化補助情報として、音声信号の波動に関する物理量（例えば周波数スペクトル情報）から少なくとも発生音を構成する各部の音の種類に関する情報を特定する解析部２５０と、符号化部２１０により符号化された符号化音声データに、解析部２５０により特定された復号化補助情報を付加する合成部２６０とを備える。この合成部２６０から出力された符号化音声データと復号化補助情報は、ＣＤ、ＤＶＤ、Ｈ／Ｄ等の記録媒体２０に記録される。なお、上記符号化音声データと復号化補助情報は、記録媒体２０内の異なる領域にそれぞれ記録されてもよい。

【００３４】一方、利用者側では、図２（ｂ）に示されたように、データ配信やＣＤ等の形態で提供された音声データが端末装置４００の外部記憶装置３０内に格納される。復号化部４１０は、キーボードや、マウス等のポインティング・デバイスなどの入力手段４６０を介して入力された利用者の指示内容に従って、外部記憶装置３０からＩ／Ｏ３１を介して読み出されたディジタルデータを所定の速度で再生可能な再生用音声データとして復号化するとともに、画像同期信号Ｄも出力する。復号化された再生用音声データはアナログデータに変換された後、スピーカー４８０から音声として出力される。

【００３５】なお、上記復号化部４１０は、外部記憶装置３０からＩ／Ｏ３１を介して読み出された音声データを読み込み、この読み出された音声データから、符号化音声データの復号化の際に参照される復号化補助情報を抽出するとともに、抽出された復号化補助情報に含まれる少なくとも発生音を構成する各部の音に関する情報を参照しながら上記符号化音声データに含まれる該発生音

を構成する各部の音ごとに音声再生に適した再生速度を、利用者が指定した再生速度情報を基準にして決定する。この復号化部４１０における符号化音声データの復号化は、該符号化音声データに含まれる発生音を構成する各部の音ごとに、上述のように決定された再生速度に相当するよう該符号化音声データの該当部分に対して伸長処理又は短縮処理を施しながら行われる。

【００３６】図３は、上述の音声データ生成装置２００における符号化部２１０の構造を示す図である。符号化部２１０は、まず、マイク等により、例えば音楽ＣＤの音響クロック４４．１ＫＨｚでサンプリングされたネイティブ・スピーカーのナチュラル・スピードの音声に相当する音声信号を取り込む。この取り込まれた音声信号は、一旦、各チャネルＣＨ＃１～ＣＨ＃８５（周波数成分）に分割するためフィルタリングされる。なお、取り込まれた音声信号の周波数範囲は７５Ｈｚ～１０，０００Ｈｚ、また、サンプリング周波数は音楽ＣＤの音響クロックに合わせて４４．１ｋＨｚ（２２．６８μｓ）である。分割するチャネル数は８５（７オクターブ＋１音）とし、各チャネル＃１～＃８５の中心周波数（中心ｆ）は平均律（１オクターブ当り１２平均律とする）の半音列になるように設定される（７７．７８Ｈｚ（Ｄ＃）～９，９６０Ｈｚ（Ｄ＃））。

【００３７】以上のように各チャネル＃１～＃８５にそれぞれ分割されたデータは、その振幅情報が２．２６８ｍｓごと（４４．１ｋＨｚサンプリングの１００データに相当、ただし１００データで１波形が形成できない場合にはデータ数を増やす）に抽出される。したがって、この実施形態では、各チャネル＃１～＃８５における振幅情報のサンプリングレート（第２周期）は４４１サンプル／ｓ（２．２６８ｍｓ）である。なお、サンプリングレートは、規則性のある周期であればよく、例えば１００データ分取り込んだ次に、１２０データ分取り込んで処理するなど、これら異なるレートで交互に処理を繰り返すような実施形態であってもよい。

【００３８】符号化部２１０は、２．２６８ｍｓごとにサンプリングされた各チャネル＃１～＃８５の振幅情報をそれぞれ１バイト（８ビット）で表現し、８５バイト（８５チャネル×１バイト）の符号化音声データａ１，ａ２，ａ３，…，ａｎを生成する。なお、符号化音声データａ１，ａ２，ａ３，…，ａｎには、該符号化音声データに相当する音声の再生時に表示される動画像との表示タイミングを制御するため、画像同期信号Ｄ（１バイト）が付加される。

【００３９】一方、図３に示された音声データ生成装置２００の解析部２５０は、上記符号化部２１０において生成された符号化音声データの復号化の際に参照される復号化補助情報を特定する。

【００４０】復号化補助情報には、取り込まれた音声信号のスペクトル情報から発生音を構成する各部の音や、強調すべき部位を示す強調位置識別情報、強調すべき周波数成分等が含まれる。

【００４１】例えば、この実施形態において、発生音を構成する各部の音は、図４に示されたように、母音部（Ｖ）と、この母音部（Ｖ）を挟んで前後に現れる子音部（図中、前子音部がC_f、後子音部がC_lで示されている）と、母音部（Ｖ）と前後の各子音部C_f、C_lとの間に現れる移行部（図中、前移行部がT_f、後移行部がT_lで示されている）と、各音声の間に現れる無音期間を示すポーズ（Ｐ）とに分類される。なお、ポーズ（Ｐ）は、再生音声の遅延の際に他の発生音を構成する各部の音と同様に延伸されたのでは、利用者の聞き取り易さを損なう可能性がある。そこで、この実施形態においてポーズ（Ｐ）は、音節の間で発生することを示す場合（Ｐ１）と、句間で発生する場合（Ｐ２）と、文間で発生する場合（Ｐ３）とにさらに分類され、それぞれ特定すべき発生音を構成する各部の音に含められている。

【００４２】上記復号化補助情報ｓ１，ｓ２，ｓ３，…，ｓｎおのおのは、符号化音声データａ１，ｓ２，ｓ３，…，ｓｎの各サンプリング間隔ごとに用意されるデータ列であって、上記発生音を構成する各部の音に関する情報として３ビット、強調位置識別情報として１ビットの計４ビット程度のデータである。また、この復号化補助情報には、非英語圏の民族には第３フォルマントのように聞き取りにくい周波数があるため、強調すべき周波数帯（特に中心周波数）の指定情報が個別に含まれてもよい。

【００４３】合成部２６０は、以上のように符号化部２１０により生成された符号化音声データａ１，ａ２，ａ３，…，ａｎに、解析部２５０により特定された復号化補助情報ｓ１，ｓ２，ｓ３，…，ｓｎを付加し、記録媒体２０にそれぞれを書き込む。なお、合成部２６０により新たに生成された音声データは、図５（ａ）～図５（ｃ）に示されたような種々の論理構造を備えることが可能である。例えば、図５（ａ）に示されたように、生成された音声データは、符号化音声データａ１，ａ２，ａ３，…，ａｎの各データごとに対応する復号化補助情報ｓ１，ｓ２，ｓ３，…，ｓｎが付加された構造であってもよい。また、図５（ｂ）に示されたように、生成された音声データは、符号化音声データａ１，ａ２，ａ３，…，ａｎと復号化補助情報ｓ１，ｓ２，ｓ３，…，ｓｎとがそれぞれ異なるグループのデータとして取り扱われる構造であってもよい。さらに、図５（ｃ）に示されたように、生成された音声データは、符号化音声データａ１，ａ２，ａ３，…，ａｎを構成する複数のグループと、復号化補助情報ｓ１，ｓ２，ｓ３，…，ｓｎの対応する複数のグループとの対により構成されてもよい。

【００４４】次に、この発明に係るデータ構造を備えた音声データの復号化及び再生を行う利用者側の端末装置４００の構造について説明する。

【００４５】図６は、端末装置４００の復号化部４１０の構造を示す図であり、図７は、図６に示された符号化部４１０におけるＰＣＭデータ生成部４１５の構造を示す図である。

【００４６】図６に示されたように、外部記憶装置３０からＩ／Ｏ３１を介して音声データが復号化部４１０に取り込まれる。なお、外部記憶装置３０に格納された音声データは、コンピュータ・ネットワークや衛星などの情報通信手段を介して配信されたり、利用者が購入したＣＤ等に格納されているデータであり、その他画像データも該外部記憶装置３０内には記録されている。また、外部記憶装置３０内に格納された音声データが圧縮されている場合には、ソフトウエア等によるデータ伸長が復号化の前処理として行われる。

【００４７】復号化部４１０では、まず抽出部４１１が、外部記憶装置３０から読み出された音声データから復号化補助情報ｓ１，ｓ２，ｓ３，…，ｓｎを抽出する。抽出された復号化補助情報ｓ１，ｓ２，ｓ３，…，ｓｎのうち、発生音を構成する各部の音に関する情報（Ｖ、C_r、C_t、T_r、T_t、Ｐ１、Ｐ２、Ｐ３）は、入力手段４６０から入力された利用者からの指示情報とともに時間係数生成部４１２に入力される。また、抽出された復号化補助情報ｓ１２，ｓ２，ｓ３，…，ｓｎのうち、強調位置識別情報は、入力処理手段４６０から入力された利用者の指示情報とともに振幅強調係数生成部４１２に入力される。さらに、抽出された復号化補助情報ｓ１２，ｓ２，ｓ３，…，ｓｎのうち、強調すべき周波数成分（中心ＣＨ）に関する情報は、入力処理手段４６０から入力された利用者の指示情報とともに強調バンドデータ生成部４１４に入力される。

【００４８】また、この実施形態では、入力手段４６０から入力される利用者の再生速度指示情報としては、図８の表に示されたように、複数の再生レベルＨ３～Ｓ６が用意されている。図８の表からも分かるように、この実施形態では、再生レベルＮを標準の再生速度（ナチュラル・スピード）とし、Ｈ３に向かうほど再生速度を早く、逆にＳ６に向かうほど再生速度を遅くするように、該ナチュラル・スピードを基準とした再生時間の比及び再生速度の倍率で指示される。

【００４９】上記時間係数生成部４１１は、図９に示されたような、再生レベル（利用者が指示）と発生音を構成する各部の音の種類との関係によって決定される再生速度倍率が予め設定された表を備えており、この表に基づいて、再生速度倍率をＰＣＭデータ生成部４１５に出力する。

【００５０】上記振幅強調係数生成部４１２は、図１０に示されたような２種類の表を備える。図１０（ａ）は、抽出部４１１によって抽出された復号化補助情報ｓ１，ｓ２，ｓ３，…，ｓｎに強調位置識別情報が含まれていない場合（強調指示がない場合）に適用される表であり、図１０（ｂ）は、復号化補助情報ｓ１，ｓ２，ｓ３，…，ｓｎに強調位置識別情報が含まれている場合（強調指示がない場合）に適用される表である。なお、これらの表に示されたパラメータは、抽出部４１１において復号化補助情報と分離された符号化音声データの各周波数成分の振幅を基準とした倍率を意味する。

【００５１】上記強調バンドデータ生成部４１４は、復号化補助情報ｓ１，ｓ２，ｓ３，…，ｓｎに強調すべき周波数帯（中心ＣＨで指定）の指示情報が含まれている場合、図１１（ａ）に示されたように、中心ＣＨに隣接する低周波数成分側５ＣＨ及び高周波数成分側５ＣＨの合計１１ＣＨについて、各周波数成分の振幅を変更するパラメータを生成する。なお、強調バンドデータ生成部４１４は、図１１（ｂ）に示されたように、再生レベルに応じた中心ＣＨの振幅倍率が予め設定された表を備えており、中心ＣＨの振幅倍率は、入力手段４６０から入力された再生速度指示情報に従って決定される。また、中心ＣＨに隣接する各ＣＨの振幅倍率は、この中心ＣＨの振幅倍率を基準にして、図１１（ａ）のように直線近似できるようにそれぞれ設定され、ＰＣＭデータ生成部４１５に出力される。

【００５２】ＰＣＭデータ生成部４１５は、図７に示されたように、各チャネルに相当する周波数成分を発生させる正弦波ジェネレータ４２２を備える。制御部４２１は、抽出部４１１からの符号化音声データから各周波数成分の振幅情報と強調バンドデータ生成部４１４からの振幅倍率データに基づいて新たに振幅係数を生成し、この生成された振幅係数を乗算器４２３において正弦波ジェネレータ４２２からのデータ（基準振幅を示す）に乗算させる。そして、得られた各周波数成分のデータが加算器４２４で加算させることにより復号化されたＰＣＭデータが得られる。さらに、制御部４２１は、時間係数生成部４１２からの再生速度倍率データに基づいて、該各符号化音声データａ１，ａ２，ａ３，…，ａｎの出力回数を調節することにより、復号化される音声データの間延びや短縮を行う。このとき、各符号化音声データａ１，ａ２，ａ３，…，ａｎごとに出力される画像同期信号Ｄの出力回数も同時に調節されることとなるため、音声データの再生側において画像の表示タイミング制御が可能になる。

【００５３】以上のようにＰＣＭデータ生成部４１５において復号化されたデータは、利用者の再生速度指示情報に従って時間軸に沿って調節された復号化データとなる。このＰＣＭデータ生成部４１５で復号化されたデータは、図１０（ａ）あるいは図１０（ｂ）のいずれかの表から振幅強調係数生成部４１２が決定した倍率パラメータと乗算器４１６において乗算される。これにより、再生用音声データが得られる。得られた再生用音声データはＤ／Ａ変換器４１７によりアナログデータに変換され、利用者が指示した再生速度の音声としてスピーカー

４８０から出力される。

【００５４】一方、端末装置４００では外部記憶装置３０から読み出された画像データの表示も可能である。図１２は、ビットマップ・ディスプレイの構造を示す図である。

【００５５】ビットマップ・ディスプレイは、１又はそれ以上のフレームを格納するメモリ４３２（ＶＲＡＭ）を備えており、描画部４３１が、外部記憶装置３０からＩ／Ｏ３２を介して読み出された画像データ（圧縮されている場合にはソフトウエア等３２によりデータ伸長される）をこれらメモリ４３２に書き込んでいく。メモリ４３２に書き込まれた画像データは１フレームごとにスイッチＳ／Ｗ４３３を介してディスプレイ４７０に表示される。なお、これら描画部４３１の書き込みタイミング及びＳ／Ｗ４３３の切り替えタイミングはタイミングコントローラ４３４により行われる。

【００５６】音声再生と画像表示とのタイミングは、この実施形態では図１３（ａ）に示されたようにＰＣＭデータ生成部４１５から出力された画像同期信号Ｄをカウントすることにより行われる。すなわち、ナチュラル・スピードでの音声再生の場合、例えば、３クロックごとにメモリ４３２のデータ書き換えを行うことにしておけば、図１３（ｂ）に示されたように、ＰＣＭデータ生成部４１５が再生速度の遅い再生用音声出たを生成する場合にも、該音声データの遅延タイミングに合ったデータ書き換えが可能になる（画像の表示タイミングを音声再生のタイミングに一致させることが可能になる）。

【００５７】すなわち、この実施形態では、メモリ４３２に格納される複数の画像データのうち１フレーム分又はそのＮ倍（正の有理数であって、Ｎは１／２や２／３であってもよい）の画像データの基準書き換え周期をＴｖ、音声のディジタル化サンプリング周期（例えば、音楽ＣＤの音響クロック）に基づいて決定された基準再生時間周期をＴａ、そして、指示された再生速度情報に基づいて発生音を構成する各部の音ごとに決定された再生時間周期をＴａ´（＞Ｔａ）とするとき、所定タイミングでの画像データ書き換え動作の終了時点からＴｖ×（（Ｔａ´／Ｔａ）－１）まで、次回の画像データ書き換え動作を休止させることを特徴としている。

【００５８】なお、音声再生と画像表示とのタイミング調整は、上述の実施形態に限定されるものではない。例えば、音声のディジタル化サンプリング周期に基づいて決定された基準再生時間周期をＴａ、そして、指示された再生速度情報に基づいて発生音を構成する各部の音ごとに決定された再生時間周期をＴａ´（＞Ｔａ）とするとき、前記メモリに格納される画像データの平均書き換え周波数は、予め指定された書き換え周波数の（Ｔａ´／Ｔａ）倍に設定されてもよい。

【００５９】

【発明の効果】以上のようにこの発明によれば、マイク等から取り込まれたり、過去に蓄積された音声信号から所定の規則に従って符号化した符号化音声データと、該符号化音声データの復号化の際に参照される発生音を構成する各部の音の種類等を含む復号化補助情報とにより音声データが構成されている。このような音声データを所定の記録媒体や配信方法により利用者に提供することにより、利用者側では任意に設定された速度の異なる複数種類の再生用音声データを復号化することができる。これにより、データ提供者から利用者へ提供すべき音声データのデータ量を低減することができ、記録媒体の記録量の節約や、データ配信時間の短縮が実現される。

【００６０】さらに、上記符号化音声データに復号化補助情報とともに画像同期信号を付加することにより、復号化された再生用音声データの再生速度に合った画像再生も可能になる。

【図面の簡単な説明】

【図１】この発明の各実施形態を概念的に説明するための図である。

【図２】（ａ）は、この発明に係る音声データの生成方法を実現する音声データ生成装置の概略構成を示すブロック図であり、（ｂ）は、この発明に係る音声データの再生方法を実現する音声データ再生装置の概略構成を示すブロック図である。

【図３】図２（ａ）に示された音声データ生成装置における符号化部の構成を示すブロック図である。

【図４】符号化された音声データの復号化に必要な復号化補助情報の一部を概念的に説明するための図である。

【図５】この発明に係る音声データのデータ構造を概念的に説明するための図である。

【図６】この発明に係る音声データの再生方法を実現する音声データ再生装置（端末装置）の構成を示すブロック図である。

【図７】図６に示された音声データ再生装置におけるＰＣＭデータ生成部の構成を示すブロック図である。

【図８】再生レベルごとに設定された、ナチュラル・スピードの再生レベルを基準とした再生時間の比率及び再生速度の倍率の一例を示す表である。

【図９】図６に示された時間係数生成部において参照される表であって、発生音を構成する各部の音の種類ごとに設定された再生速度の一例を、ナチュラル・スピードの再生レベルを基準とした倍率で示した表である。

【図１０】図６に示された振幅強調係数生成部において参照される表であって、発生音を構成する各部の音の種類ごとに設定される振幅の一例を、ナチュラル・スピードの再生レベルを基準とした倍率で示した表である。

【図１１】（ａ）は、図６に示された強調バンドデータ生成部において参照される表であって、指示された周波数バンドデータの編集動作を説明するための図であり、（ｂ）は、指示された周波数バンド（中心ＣＨ）の振幅を、ナチュラル・スピードの再生レベルを基準とした倍

率で示した表である。

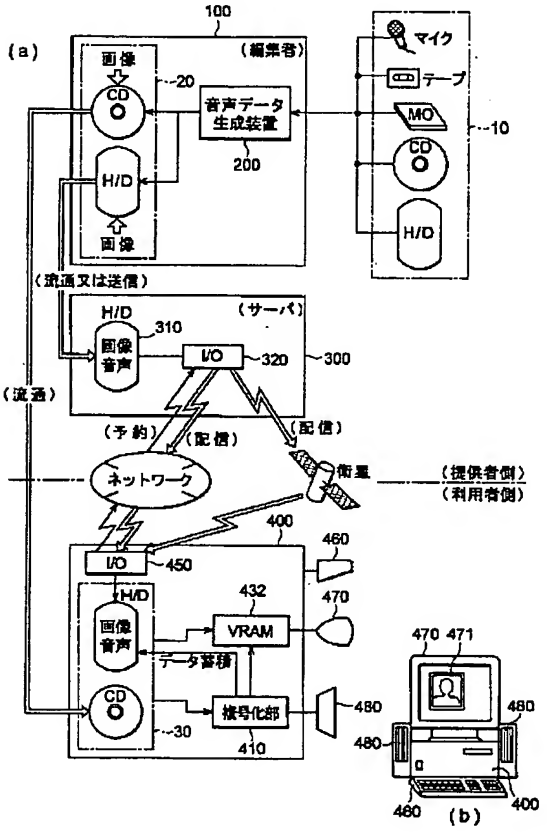
【図１２】図６に示された音声データ再生装置（端末装置）による音声再生に同期して画像データを表示する表示装置の構成を示す図である。

【図１３】音声再生動作に同期した画像表示タイミングを説明するためのタイムチャートである。

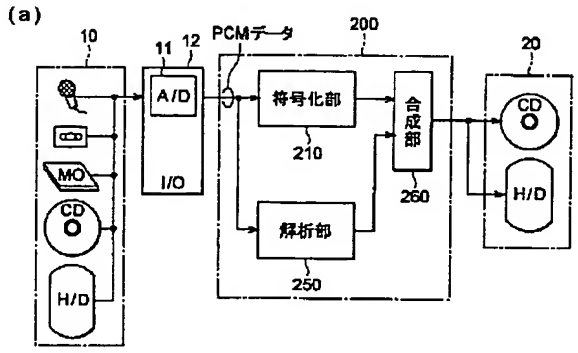
【符号の説明】

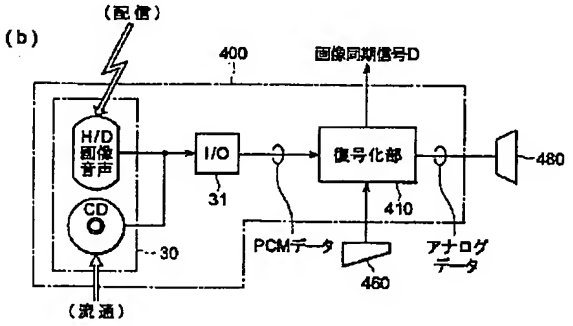
１０…情報源、２０、３０、３１０…記録媒体、２００…音声データ生成装置、２１０…符号化部、２５０…解析部、２６０…合成部、４００…端末装置（ＰＣ）、４１０…復号化部。

【図１】

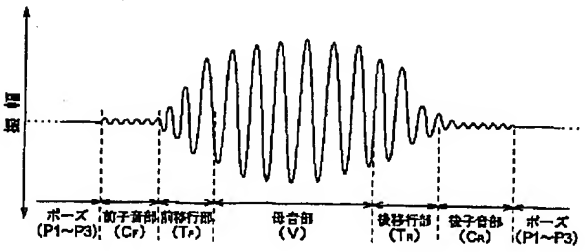


【図２】

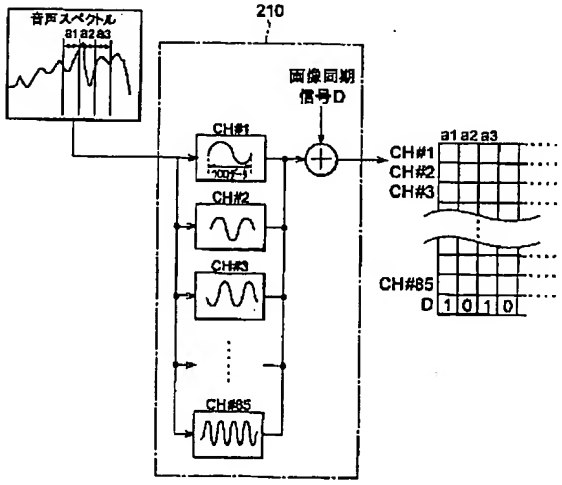




【図４】



【図３】

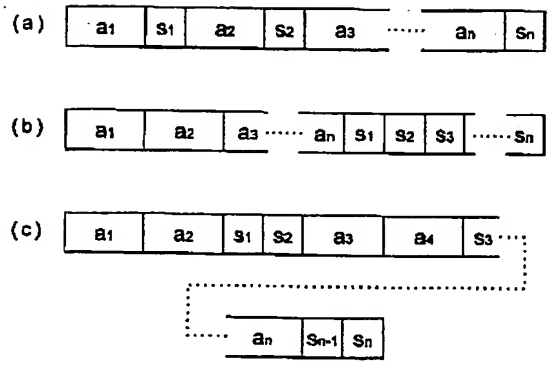


【図８】

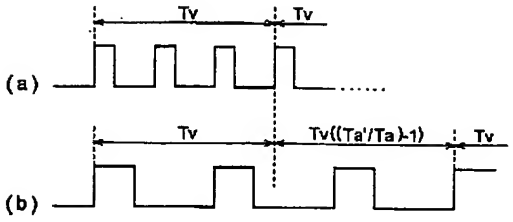
再生レベル	H3	H2	N	S2	S3	S4	S5	S6
再生時間比	0.64	0.80	1.00	1.25	1.56	1.95	2.44	3.05
再生速度倍率	1.56	1.25	1.00	0.80	0.64	0.51	0.41	0.33

速い←　　→遅い

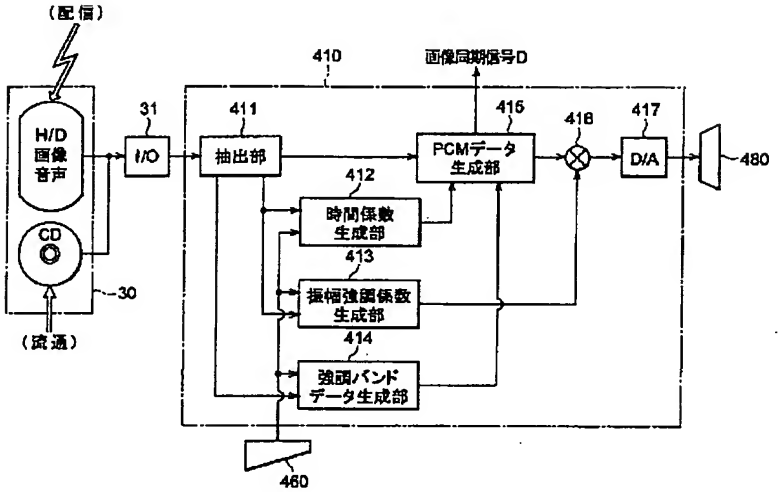
【図５】



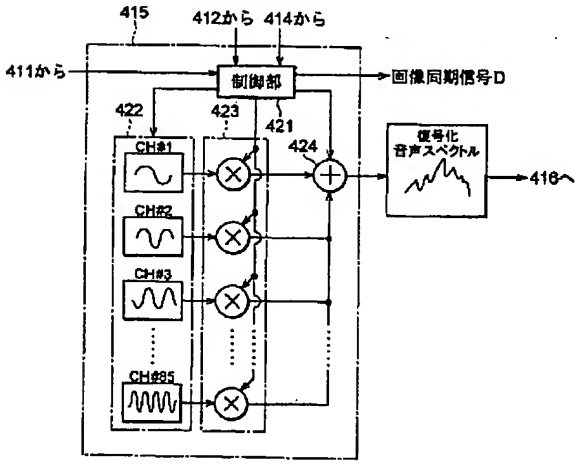
【図１３】



【図６】



【図７】



【図９】

再生速度倍率 ＼ 発生音	CF	TF	V	TR	CR	P1	P2	P3
0.33	0.64	0.64	0.25	0.64	0.64	0.33	0.80	0.80
0.41	0.64	0.64	0.33	0.64	0.64	0.41	0.80	0.80
0.51	0.64	0.64	0.41	0.64	0.64	0.51	0.80	0.80
0.64	0.64	0.64	0.51	0.64	0.64	0.64	0.80	0.80
0.80	0.80	0.80	0.64	0.80	0.80	0.80	0.80	1.00
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
1.25	1.10	1.10	1.30	1.10	1.10	1.10	1.05	1.00
1.56	1.25	1.25	1.70	1.25	1.25	1.25	1.10	1.00

遅い ↑　↓ 速い

【図１０】

(a)

再生速度倍率＼発生音	CF	TF	V	TR	CR
0.33	2.00	1.50	1.00	1.50	2.00
0.41	1.75	1.32	1.00	1.32	1.75
0.51	1.52	1.23	1.00	1.23	1.52
0.64	1.32	1.15	1.00	1.15	1.32
0.80	1.15	1.07	1.00	1.07	1.15
1.00	1.00	1.00	1.00	1.00	1.00
1.25	1.15	1.10	1.00	1.10	1.15
1.56	1.16	1.10	1.00	1.10	1.15

遅い ↑　↓ 速い

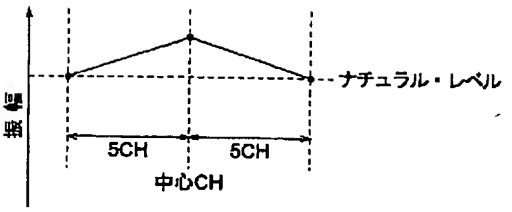
(b)

再生速度倍率＼発生音	CF	TF	V	TR	CR
0.33	2.40	1.95	1.30	1.95	2.40
0.41	2.28	1.72	1.30	1.72	2.28
0.51	2.00	1.60	1.30	1.60	2.00
0.64	1.72	1.50	1.30	1.50	1.72
0.80	1.31	1.22	1.14	1.22	1.31
1.00	1.00	1.00	1.00	1.00	1.00
1.25	1.31	1.22	1.14	1.22	1.31
1.56	1.31	1.22	1.14	1.22	1.31

遅い ↑　↓ 速い

【図１１】

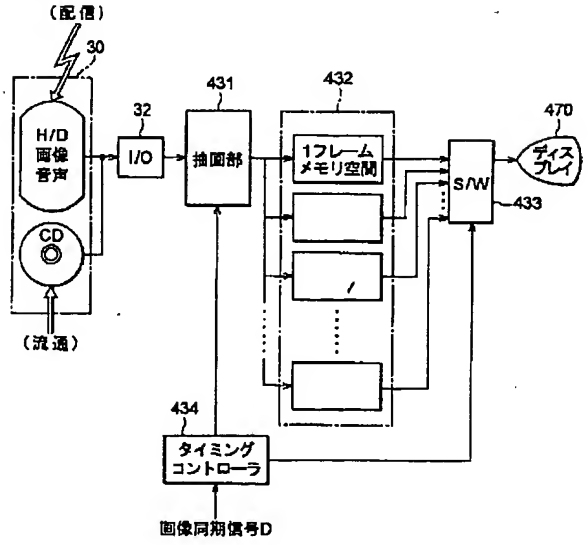
(a)



(b)

再生速度倍率	中心CH
0.33	1.20
0.41	1.20
0.51	1.20
0.64	1.12
0.80	1.06
1.00	1.00
1.25	1.00
1.56	1.00

【図１２】



フロントページの続き

(51)Int.Cl.'	識別記号	FI	テーマコード'(参考)
// Ｈ０４Ｎ　7/173	６３０	Ｈ０４Ｎ　5/92	Ｈ　５Ｄ０８０
			９Ａ００１

Ｆターム(参考)　2C028 AA03 BA01 BB04 BB07 BD03
CA11 CA13 CB03
5C053 FA22 FA24 FA29 GA11 GB11
JA07 JA12 KA04 LA11 LA15
5C064 BA01 BB07 BC16 BC23 BC25
BD02 BD07 BD08
5D044 AB05 BC01 BC04 CC01 CC04
CC09 DE17 DE49 EF05 FG18
GL50
5D045 DA20
5D080 BA01 BA03 FA39 GA16 GA22
9A001 EE01 HH15 JJ19 KK09